

**ASPEKTE DER ZEUGENKOMPETENZ und
VALIDIERUNG DER KRITERIENORIENTIERTEN
AUSSAGEANALYSE von
JUGENDLICHEN MIT INTELLIGENZMINDERUNG**

*Längsschnittliche Untersuchungen zur Gedächtnisleistung und
Quellendifferenzierung sowie Analyse der aussageimmanenten
Qualitätsmerkmale*

Inaugural – Dissertation
zur Erlangung der Doktorwürde (Dr. phil.)
der Philosophischen Fakultät II
(Psychologie, Pädagogik und Sportwissenschaft)
der Universität Regensburg



vorgelegt von
Oliver Berger aus
Regensburg
2005

Erstgutachter: Prof. Dr. Alf C. Zimmer

Zweigutachter: Prof. Dr. Alexander Thomas

DANKSAGUNGEN

An dieser Stelle möchte ich mich bei all den Menschen bedanken, die in welcher Form auch immer Anteil an dieser Dissertation haben und mich in fachlicher oder persönlicher Hinsicht unterstützt haben.

Zuerst gilt mein Dank meinem Doktorvater, Herrn Prof. Dr. Alf Zimmer, der die Betreuung und Begutachtung der vorliegenden Arbeit übernahm und mir vielseitige fachliche Anregungen und Tipps zukommen ließ. Trotz seiner knappen Zeit und Position als Rektor bestand fast jederzeit die Möglichkeit und das Angebot einer fachlichen Aussprache.

Ebenfalls vielen Dank an Herrn Prof. Dr. Alexander Thomas, der sich bereit erklärt hat, die Zweitbegutachtung dieser Dissertation zu übernehmen und mir immer sehr kooperativ begegnete!

Im Besonderen möchte ich mich bei Frau Dr. Sandra Loohs bedanken, die nicht nur meine direkte Ansprechpartnerin und Unterstützung bei der Planung und Ideenumsetzung war, bereits meine Diplomarbeit betreute und mir auch bei der Dissertation als engagierte Diskussionspartnerin zur Seite stand, sondern über die vielen Jahre unserer Bekanntschaft und Zusammenarbeit zu einer fachlich wie persönlich hoch geschätzten Vertrauten wurde. In erster Linie durch sie wurde vor Jahren mein Interesse an der forensischen und Aussagepsychologie geweckt. Sandra, vielen Dank für alles!

Im Rahmen der Durchführung der Untersuchung (Datenerhebung) gilt mein herzlichster Dank allen Beteiligten und Verantwortlichen des Berufsbildungswerk St. Franziskus in Abensberg. Im einzelnen Herrn Direktor Dr. Peter Schopf, der mit seiner Zustimmung überhaupt erst die Erhebung der Interviews ermöglichte. Zum zweiten Herrn Dipl.-Psych. Hans Hinterobermaier für die große informatorische Unterstützung und Hilfestellung in den unterschiedlichsten Bereichen. Ganz besonders möchte ich mich bei Frau Evi Rosenmeier bedanken, die eine unglaublich große organisatorische Meisterleistung vollbrachte und die wichtigste Unterstützung bei der konkreten Organisation und im logistischen Ablauf der vielen Interviews war. Ohne sie wäre die reibungslose Durchführung der Befragungen über einige Wochen hinweg und hinsichtlich der komplexen Strukturen einer solchen Institution so nicht möglich gewesen. Vielen Dank!

Natürlich sei auch allen Jugendlichen der Einrichtung bzw. deren Erziehungsberechtigten gedankt, die sich bereit erklärt hatten, an den oftmals mehrfach stattfindenden Terminen teilzunehmen.

In Bezug auf die Rahmenbedingungen der Promotion gilt mein Dank dem Freistaat Bayern bzw. der Universität Regensburg, die meine Promotion im Rahmen der Graduiertenförderung durch ein Stipendium zur Förderung des wissenschaftlichen Nachwuchses finanziell unterstützte, für das in mich gesetzte Vertrauen.

Im privaten Bereich möchte ich mich bei allen Freunden und Bekannten bedanken, die mich in den vergangenen zweieinhalb Jahren unterstützt haben, insbesondere in den letzten, besonders arbeitsreichen Monaten und mich trotz Widerstandes meinerseits auch mal vom Schreibtisch weggezerrt haben... oder es zumindest versucht haben. Danke.

Zu guter Letzt möchte ich mich ganz besonders bei meiner lieben Mama bedanken, die mich immer und zu jeder Zeit motivierte, geduldig meinen Stress und die manchmal aufgekommene schlechte Laune ertrug und mir auch in privat schweren Zeiten während dieser Promotion, immer das Gefühl gab, es geht trotzdem weiter und die Entscheidung, diesen Weg so zu gehen, ist die richtige. Vielen Dank! Diese Dissertation sei Dir gewidmet!

Oliver Berger

Mai 2005

VORWORT

Im Rahmen der seit Mitte der 1970er Jahre experimentell untersuchten forensischen Psychologie und der Psychologie der Zeugenaussage liegt mittlerweile eine enorme Bandbreite von Untersuchungen vor, die sich unter anderem mit bei Zeugenaussagen auftretenden Unzuverlässigkeiten (1), Validierungsmethoden zur Einschätzung der Glaubhaftigkeit von Zeugenaussagen (2), Untersuchungen zur Personenbeschreibung und – identifikation (3) bis hin zur Entwicklung spezifischer Befragungsmethoden, die einen möglichst vollständigen und fehlerfreien Abruf des Erlebten unterstützen sollen (4), beschäftigen. Alle genannten Aspekte sind in der vorliegenden Arbeit von Bedeutung, jedoch insbesondere die Punkte (1) und (2).

Die Begutachtung von Zeugenaussagen im Sinne der so genannten Glaubhaftigkeitsbegutachtung sowie die Analyse von Kompetenzmerkmalen *von Menschen mit kognitiven und intellektuellen Einschränkungen* stellen den forensisch-angewandten Hintergrund und spezifischen Aspekt der vorliegenden Dissertation dar.

Neben dem Aspekt der Aussagezuverlässigkeit sind hinsichtlich der Begutachtung der Glaubhaftigkeit von Zeugenaussagen ebenso besonders die Konstrukte der Aussagequalität und der Aussagetüchtigkeit relevant.

Bezüglich der Qualität einer Zeugenaussage kann aufgrund der inzwischen gut abgesicherten Forschungsergebnisse festgestellt werden, dass sich Aussagen, die sich auf ein reales, tatsächlich erlebtes oder beobachtetes Ereignis beziehen, in bestimmtem inhaltlichen Qualitätsmerkmalen von Schilderungen über entsprechende Ereignisse ohne diesen so genannten Erlebnishintergrund, unterscheiden. Trotzdem diese Grundannahme für Kinder und Erwachsene als allgemein bestätigt angesehen wird und von einigen Forschern nun deshalb schon seit längerem daraufhin hingewiesen wurde, die Forschung nun auf spezielle Personengruppen, wie etwa Menschen mit intellektuellen Einschränkungen oder sonstigen Behinderungen, auszuweiten, anzuwenden bzw. durch empirische Studien Erkenntnisse diesbezüglich zu gewinnen, blieb dieser Vorschlag in machen Aspekten offenkundig weitestgehend unberücksichtigt.

Auch hinsichtlich des Aspekts der Aussagetüchtigkeit, der sich vor allem auf die kognitiven-funktionalen Leistungsvoraussetzungen einer Person bezieht, stellt sich in dem genannten Zusammenhang die Frage, was von einer Zeugenaussage einer Person mit Intelligenzminderung überhaupt erwartet werden kann, in welchen Bereichen und in welchem Ausmaß Defizite vorhanden sind und welche Maßstäbe (der Kompetenz oder Glaubhaftigkeit) hier angelegt werden müssen und können.

Der beschriebene zum Teil defizitäre Forschungsstand gestaltet sich umso tragischer, da sich zudem die Meldungen und Befunde häufen, dass Menschen mit Behinderungen immer öfter Opfer von Straftaten gegen die sexuelle Selbstbestimmung werden und es in der Folge (leider in viel zu wenigen Fällen) zu einer strafrechtlichen Verfolgung der Täter kommt, in deren Verlauf die Opfer(-zeugen) oft von mehreren unterschiedlichen Personen und Institutionen zum inkriminierten Geschehen befragt werden. Hierbei sind im Fall von Personen mit Behinderungen auf Seiten der Befragenden *aufgrund der Behinderung* oft Unwissenheit, Verunsicherung sowie auch Zweifel an der Glaubhaftigkeit der Aussage und auch der Glaubwürdigkeit der Person selbst festzustellen. Dies mag unter anderem den Hintergrund haben, dass die Phänomene Intelligenzminderung oder Lernbehinderung allein bezüglich ihrer Terminologie wie auch der Komplexität und Individualität nicht ganz problemlos erscheinen und dementsprechend kritisch zu betrachten sind.

Es erscheint daher wohl aus Sicht der psychologischen Forschung, aber auch aus juristischer Perspektive angeraten, sich verstärkt mit den Kompetenzmerkmalen dieser Personen-Gruppe auseinander zu setzen.

Neben den rein gedächtnispsychologischen Aspekten und Prozessen spielen hierbei auch soziale und motivationale Bedingungen eine entscheidende Rolle bei der Einschätzung und Feststellung der Erinnerungsleistung eines Zeugen.

Konkret gilt es aufzuzeigen, dass sich eine Zeugenaussage als veränderliches, komplexes, von vielen subjektiven wie objektiven Faktoren sowie externalen und internalen Prozessen beeinflusstes Gebilde darstellt.

Auf dem Hintergrund eines notwendigen Zeugen- und Opferschutzes, insbesondere von Menschen mit Intelligenzminderung, ist es Aufgabe und Anspruch dieser Arbeit „etwas Licht ins Dunkle der Phänomene ‘Lernbehinderung’ und ‘Menschen mit kognitiven Defiziten als Zeugen’ zu bringen“ und das Vorurteil der „pauschal unbrauchbaren“ Zeugenaussage solcher Personen zu widerlegen.

Die vorliegende Dissertation wird und kann die angesprochenen Themen sicher nicht erschöpfend darstellen und/oder klären können, soll in dieser Hinsicht aber zumindest als Anstoß für weitere spezifische (d.h. auf spezielle Personengruppen bezogene und ökologisch valide), dringend notwendige, auch interdisziplinäre Forschungsarbeiten dienen.

INHALT

Danksagungen	i
Vorwort.....	iii
Inhaltsverzeichnis	v

1 Einleitung –**Theoretische Grundlagen und empirischer Hintergrund..... 1**

1.1 Allgemeine Aspekte zur Gedächtnispsychologie	1
1.1.1 Gedächtniskomponenten und Gedächtnisprozesse.....	1
1.1.2 Relevante Bereiche des Gedächtnisses für Zeugenaussagen.....	5
1.1.3 Alltagsgedächtnis	7
1.1.3.1 Autobiographisches und Augenzeugengedächtnis	8
1.1.3.2 Der Begriff des Augenzeugen.....	11
1.1.4 Generelle Einflussfaktoren auf die Gedächtnisleistung	13
1.1.5 Das Konzept der Realitätsüberwachung (reality monitoring)	17
1.1.6 Das Konzept der Quellendifferenzierung oder Quellenkontrolle (source monitoring)	19
1.2 Psychologie der Zeugenaussage	21
1.2.1 Fehler und Unzuverlässigkeiten bei Zeugenaussagen – Modelle und Ursachen.....	21
1.2.1.1 Einflussfaktoren und Fehler bei Quellenidentifikation und Quellenzuordnung	23
1.2.1.2 Formen und Ursachen suggestiver und nachfolgender Beeinflussung	32
1.2.1.3 sonstige Aspekte	43

1.2.2	Aspekte der Glaubhaftigkeitsbegutachtung	51
1.2.2.1	Die aussagepsychologische Konstrukt-Trias	53
1.2.2.2	Integrative Kriteriologie nach Steller und Köhnken (1989)	58
1.2.2.3	Integrierende Glaubhaftigkeitsbeurteilung.....	73
1.2.2.4	Empirische Befundlage zur Validierung der kriterienorientierten Aussageanalyse (CBCA)	75
1.2.3	Juristische Aspekte	79
1.2.3.1	Das BGH-Urteil: Methodische und formale Standards der Glaubhaftigkeitsbegutachtung	79
1.2.3.2	Prozessrechtliche Stellung des Zeugen und des Sachverständigen	81
1.3	Intelligenzminderung und Lernbehinderung.....	83
1.3.1	Terminologie	83
1.3.1.1	Definitionen nach ICD-10 bzw. WHO	83
1.3.1.2	Konzepte, Modelle und begriffliche Abgrenzungen: Lernbehinderung - Lernschwäche - Lernbeeinträchtigung - geistige Behinderung	85
1.3.1.3	Analogien und Differenzen zum angloamerikanischen Sprachraum	96
1.3.2	Sexueller Missbrauch und sexuelle Gewalt gegen Menschen mit Intelligenzminderung	98
1.3.2.1	Empirische Befunde und Untersuchungen zu Prävalenz und Inzidenz	99
1.3.2.2	erhöhte Vulnerabilität von Menschen mit kognitiven Einschränkungen – Analyse von Bedingungs- und Risikofaktoren und Erklärungsansätze	104
1.3.2.3	Aspekte zum Täterverhalten	111
1.3.2.4	Rechtslage und Strafrechtsnormen.....	113
1.4	Menschen mit Intelligenzminderung als Zeugen.....	115
1.4.1	Beurteilung der Aussagetüchtigkeit bei Menschen mit Intelligenzminderung ...	116

1.4.1.1	Aspekte der Zeugenkompetenz von Menschen mit kognitiven Einschränkungen.....	117
1.4.1.2	Gedächtnisleistungen - Quellendifferenzierung und Realitätsüberwachung bei Intelligenzminderung	122
1.4.1.3	Gedächtnisleistungen - Suggestibilität und Beeinflussbarkeit bei Intelligenzminderung	129
1.4.2	Interviewtechniken und Implikationen für eine optimale Befragungssituation...	137
2	Fragestellung.....	142
2.1	Fragen zur Untersuchung der Erinnerungsleistung und dem source- bzw. reality monitoring.....	142
2.1.1	Fragen zu den Unterschieden zwischen externaler 'source monitoring' - Gruppe und 'reality monitoring' - Gruppe	142
2.1.2	Fragen zur Veränderung der Gedächtnisleistung über die Zeit.....	143
2.1.3	Fragen zum Zeitpunkt der Befragung	143
2.1.4	Fragen zu Haupt- und Interaktionseffekten.....	144
2.1.5	sonstige Fragen.....	145
2.2	Fragen zur Glaubhaftigkeitsuntersuchung.....	145
2.2.1	Fragen zum allgemeinen Differenzierungspotential der Aussagemerkmale.....	145
2.2.2	Fragen hinsichtlich verschiedener Befragungszeitpunkte und des Delays	145
2.2.3	Fragen hinsichtlich der gutachterlichen, forensischen, polizeilichen und richterlichen Praxis	146
2.2.4	Fragen hinsichtlich der Anwendbarkeit auf allgemeinspsychologische Wahrnehmungs- und Gedächtnisprozesse.....	146

3 Methoden	147
3.1 Untersuchung zur Erinnerungsleistung und dem ‘source- bzw. reality monitoring’	147
3.1.1 Versuchspersonen	147
3.1.2 Versuchsmaterial.....	149
3.1.2.1 Der Inhalt der Dia-Serien	150
3.1.2.1.1 Dia-Serie 1 – PKW-PKW	150
3.1.2.1.2 Dia-Serie 2 – PKW-Fahrrad	150
3.1.2.2 Die selbst erfundene Unfall-Serie – analog Dia-Serie ‘PKW-Fahrrad’	151
3.1.2.3 Präsentation des Versuchsmaterials und Instruktion der Versuchspersonen	151
3.1.2.3.1 externale ‘source monitoring’ - Gruppe	151
3.1.2.3.2 ‘reality monitoring’ - Gruppe	152
3.1.3 Experimentalplan.....	152
3.1.4 Durchführung der Befragungen	154
3.1.4.1 Zeitlicher Abstand der Befragungen zur Stimuluspräsentation.....	154
3.1.4.2 Ablauf der Befragungen.....	155
3.1.5 Auswertung der Befragungen	156
3.1.5.1 Transkription der Interviews.....	156
3.1.5.2 Auswertung der Transkripte.....	156
3.1.6 Variablenübersicht.....	159
3.1.7 Statistische Auswertung	160
3.1.7.1 Signifikanztests.....	160
3.1.7.2 Der Effektstärkenindex d beim Vergleich zweier Stichprobenmittelwerte	161

3.2	Untersuchung zur Glaubhaftigkeit von Zeugenaussagen	162
3.2.1	Versuchspersonen	162
3.2.2	Grundlage für die Befragungen und Instruktion der Versuchspersonen.....	163
3.2.2.1	Simulationsstudie vs. Feldstudie.....	163
3.2.2.2	Grundvariablen (Kriterien) der und Themen für die Geschichten	163
3.2.2.3	Wahl der Themen und Vorgehensweise	164
3.2.3	Experimentalplan.....	165
3.2.3.1	Blindversuch.....	165
3.2.3.2	Untersuchungsdesign.....	166
3.2.4	Durchführung der Befragungen	166
3.2.4.1	Befragungszeitpunkte	166
3.2.4.2	Ablauf der Befragungen.....	167
3.2.5	Auswertung der Befragungen	168
3.2.5.1	Transkription und Rating der durchgeführten Interviews	168
3.2.5.2	Zur Problematik der Auswertung, des Ratingverfahrens und Skalenniveaus verbaler Berichte	168
3.2.5.3	Das Prinzip der Aggregation	170
3.2.6	Variablenübersicht.....	171
3.2.7	Statistische Auswertung	173
3.2.7.1	Signifikanztests.....	173
3.2.7.2	Der Effektstärkenindex d beim Vergleich zweier Stichprobenmittelwerte ...	173
3.2.7.3	Übersicht über die angewandten statistischen Verfahren	174
4	Ergebnisse	175
4.1	Ergebnisse zur Erinnerungsleistung und dem 'source- bzw. reality monitoring'	175
4.1.1	Überprüfung der (intervallskalierten) Variablen auf	

Verteilungsform und Varianzhomogenität	175
4.1.1.1 Überprüfung der Verteilungsform.....	175
4.1.1.2 Überprüfung der Varianzhomogenität	177
4.1.2 Ergebnisse der Unterschiede zwischen externaler 'source monitoring' - und 'reality monitoring' - Gruppe	178
4.1.2.1 Ergebnisse zum ersten Befragungszeitpunkt T1.....	178
4.1.2.2 Ergebnisse zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging	186
4.1.2.3 Ergebnisse zum zweiten Befragungszeitpunkt NT2 ohne vorausgehende Befragung	194
4.1.3 Ergebnisse der Veränderungen der Gedächtnisleistung über die Zeit	206
4.1.3.1 Ergebnisse für die externe 'source monitoring' – Gruppe	206
4.1.3.2 Ergebnisse für die 'reality monitoring – Gruppe'	213
4.1.4 Ergebnisse zu den Befragungszeitpunkten.....	220
4.1.4.1 Befragungszeitpunkt T1 vs. NT2 (erste Befragung nach einer bzw. acht Wochen)	220
4.1.4.1.1 Ergebnisse für die externe 'source monitoring' – Gruppe	220
4.1.4.1.2 Ergebnisse für die 'reality monitoring' – Gruppe	227
4.1.4.2 Befragungszeitpunkt T2 vs. NT2 (vorausgehende Befragung bei T1 vs. ohne vorausgehende Befragung - jeweils nach acht Wochen	234
4.1.4.2.1 Ergebnisse für die externe 'source monitoring' – Gruppe	234
4.1.4.2.2 Ergebnisse für die 'reality monitoring' – Gruppe	241
4.1.5 Ergebnisse zu den Interaktionseffekten.....	249
4.2 Ergebnisse zur Glaubhaftigkeit von Zeugenaussagen (Validierungsuntersuchung)	257
4.2.1 Überprüfung auf Verteilungsform und Varianzhomogenität	258

4.2.2	Ergebnisse zu den einzelnen Aussagemerkmalen	260
4.2.2.1	Logische Konsistenz.....	260
4.2.2.2	Widerspruchsfreie, ungeordnete Darstellung	262
4.2.2.3	Quantitativer Detailreichtum.....	264
4.2.2.4	Kontextuelle Einbettung	266
4.2.2.5	Schilderung nonverbaler Aktionen	268
4.2.2.6	Wiedergabe von Gesprächen	270
4.2.2.7	Schilderung von Komplikationen im Handlungsverlauf.....	272
4.2.2.8	Ausgefallene Details	274
4.2.2.9	Nebensächliche Details	276
4.2.2.10	Schilderung eigener psychischer oder physischer Vorgänge	278
4.2.2.11	Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	280
4.2.2.12	Spontane Verbesserungen der eigenen Aussage	282
4.2.2.13	Zugeben von Erinnerungslücken	284
4.2.2.14	Einwände gegen die Richtigkeit der eigenen Aussage.....	286
4.2.2.15	Selbstbelastungen	288
4.2.2.16	Inschutznahme des Täters oder Beteiligten	290
4.2.2.17	Deliktspezifische Aussageelemente.....	292
4.2.3	Ergebnisse zu den Merkmalskategorien und zu allen Aussagemerkmalen als Gesamtinstrument	294
4.2.3.1	Allgemeine Merkmale	294
4.2.3.2	Spezielle Inhalte	295
4.2.3.3	Inhaltliche Besonderheiten	295
4.2.3.4	Motivationsbezogene Inhalte	296
4.2.3.5	Deliktspezifische Inhalte	296
4.2.3.6	Aussagemerkmale 'total'	297

4.2.4	Tabellarische Kurzzusammenfassung aller Ergebnisse.....	298
4.2.4.1	Ergebnisse zu den einzelnen Aussagemerkmalen.....	298
4.2.4.2	Ergebnisse zu den Merkmalskategorien und für alle Aussagemerkmale zusammen.....	301
4.2.5	Ergebnisse zur Glaubhaftigkeitseinschätzung	302
4.2.5.1	Glaubhaftigkeitseinschätzung zu beiden Befragungszeitpunkten.....	303
4.2.5.2	Glaubhaftigkeitseinschätzung zum ersten Befragungszeitpunkt T1	304
4.2.5.3	Glaubhaftigkeitseinschätzung zum zweiten Befragungszeitpunkt T2	305
4.2.6	Klassifikation der Glaubhaftigkeit anhand der erhobenen Aussagemerkmale mittels Diskriminanzanalyse	306
4.2.6.1	Analyse der Gruppenunterschiede und Prüfung der Trennfähigkeit der Indikatoren zu beiden Befragungszeitpunkten	306
4.2.6.2	Die diskriminatorisch bedeutsamen Aussagemerkmale zu T1 und T2	312
4.2.6.3	Die Prüfung der Diskriminanzfunktion.....	313
4.2.6.4	Der kritische Diskriminanzwert.....	314
4.2.6.5	Klassifizierungsergebnisse	315
4.3	Zusammenfassung der Ergebnisse.....	317
4.3.1	Erinnerungsleistung und 'source- bzw. reality monitoring'	317
4.3.1.1	Ergebnisse der Unterschiede zwischen den Versuchsgruppen.....	317
4.3.1.2	Ergebnisse der Veränderung der Gedächtnisleistung über die Zeit	322
4.3.1.3	Ergebnisse der Gedächtnisleistungen im Vergleich der einzelnen Befragungszeitpunkte zueinander	324
4.3.1.4	Ergebnisse zu den Haupt- und Interaktionseffekten.....	330
4.3.2	Glaubhaftigkeit von Zeugenaussagen	331
4.3.2.1	Ergebnisse zu den einzelnen Aussagemerkmalen.....	331
4.3.2.2	Ergebnisse zu den Merkmalskategorien und zu allen Aussagemerkmalen als Gesamtinstrument.....	332

4.3.2.3	Ergebnisse zur Glaubhaftigkeitseinschätzung	334
4.3.2.4	Ergebnisse zur Klassifikation	335
5	Diskussion	336
5.1	Diskussion zu den Befunden der Erinnerungsleistung und dem ‘source- bzw. reality monitoring’	336
5.1.1	Diskussion zu den Ergebnissen zwischen den Versuchsgruppen.....	337
5.1.2	Diskussion zu den Ergebnissen der Veränderung der Gedächtnisleistung über die Zeit	345
5.1.3	Diskussion zu den Ergebnissen der Gedächtnisleistungen zu den einzelnen Befragungszeitpunkten	352
5.1.4	Diskussion zu den Interaktionseffekten	362
5.2	Diskussion zu den Befunden der Glaubhaftigkeit von Zeugenaussagen (Validierungsuntersuchung).....	364
5.2.1	Diskussion zu den Befunden der einzelnen Aussagemerkmale.....	364
5.2.2	Diskussion zu den Befunden über alle Aussagemerkmale hinweg und zu den Merkmalskategorien.....	378
5.2.3	Diskussion zu den Befunden der diskriminatorisch bedeutsamen Aussagemerkmale, den Klassifizierungsergebnissen und Glaubhaftigkeitseinschätzungen	383
5.3	Zusammenfassende und sonstige Aspekte	390
5.4	Ökologische Validität sowie Implikationen und Ausblick für zukünftige Forschungsarbeiten und die forensische Praxis	400

Zusammenfassung	408
Literaturverzeichnis.....	415
Anhang	445
Abbildungs- und Tabellenverzeichnis	486

1. Einleitung – Theoretische Grundlagen und empirischer Hintergrund

1.1 Allgemeine Aspekte zur Gedächtnispsychologie

1.1.1 Gedächtniskomponenten und Gedächtnisprozesse

Nach ATKINSON und SHIFFRIN (1968) werden das so genannte sensorische Register, der Kurz- und der Langzeitspeicher als distinkte Speicherstrukturen des Gedächtnisses angesehen. Zwischen diesen Speichereinheiten soll der Informationsfluss linear progredient in verschiedenen Stadien verlaufen sowie unterschiedlichen spezifischen Kontroll- bzw. Verarbeitungsprozessen, wie z. B. selektiver Aufmerksamkeit oder spezifischen Gedächtnisstrategien unterliegen (vgl. SCHLEIDER, 1993).

Das *sensorische Register* registriert Informationen, die je nach Modalität (visuell, sensorisch oder echoisch) rezipiert und transformiert werden, wobei es zu modalitätsspezifischen Interferenzen kommen kann. Auf dieser Stufe der Verarbeitung ist die Repräsentation kaum von Erfahrungen, selektiver Aufmerksamkeit oder Stimulusmerkmalen abhängig. Es erfolgt eine reiznahe und passive Verarbeitung bei der die Bedeutung der Information noch nicht identifiziert wird. Die Kapazität dieses Speichers ist als sehr groß und wichtig einzuschätzen, jedoch zerfällt die Information schon nach wenigen Millisekunden.

Der *Kurzzeitspeicher (KZS)* ist als aktiver Teil des Langzeitspeichers anzusehen, in dem nur eine begrenzte Menge an Information zur weiteren Verarbeitung bereitgehalten werden kann und die ohne aktive Kontrollmaßnahmen und unter Mitwirkung von Interferenzprozessen nach ca. 30 Sekunden zerfällt.

Ein anderes Konzept des KZS ist das des *Arbeitsspeichers* (*'working memory'*) nach BADDELEY (1986). Es beschreibt den KZS als aktives System, das für Speicherung bzw. Aktivierung aufgabenrelevanter Informationen und deren Transformation zuständig ist, wobei die Verarbeitungskapazität je nach Anzahl der Speicherplätze, die zur passiven Speicherung benötigt werden, mehr oder weniger einschränkt. BRAINERD (1983) geht davon aus, dass die Kapazität des Arbeitsgedächtnisses kontextfrei, also nicht primär unter der Kontrolle der zu lösenden Aufgabe stehe. Nur innerhalb einer begrenzten Gesamtkapazität und eines zeitlich limitierten Akts käme es zum Bereithalten von Informationen und zum Abarbeiten von Verarbeitungsroutinen, wobei jeweils die eine Vorgehensweise auf Kosten der anderen gehe. So wird das Arbeitsgedächtnis als aktiver Teil der menschlichen Informationsverarbeitung, gleichzeitig aber auch als kapazitätsbeschränkendes System verstanden.

Der *Langzeitspeicher (LZS)* ist ein scheinbar unbegrenztes und permanentes System, das das „Wissen über die Welt“ enthält. Diese sog. epistemische Struktur unterteilt sich in:

- (1) das semantische Gedächtnis mit prozeduralem, deklarativem und perzeptuellem Wissen und
- (2) das episodische Gedächtnis, wo die Repräsentation vor allem perzeptuell und/oder semantisch erfolgt.

Im Folgenden sollen nur kurz einige Erklärungsansätze zur Entwicklung der Gedächtniskapazität skizziert werden (vgl. ROTH, 1997):

- (1) *Theorie konstruktiver Operatoren*: Diese Theorie ist den strukturellen Ansätzen des KZS zuzuordnen. Demnach sind intellektuelle Leistungen durch zwei Komponenten bestimmt:
 - (a) das Repertoire verfügbarer figurativer, operativer und exekutiver Schemata (vergleichbar mit Wissens-, Strategie- und Metakognitionskomponenten)
 - (b) die strukturelle Kapazität (mental-space)
- (2) *Neuformulierung von (1) durch CASE*: Der Gesamtspeicher ist in zwei Subsysteme eingeteilt:
 - (a) einen passiven Speicherraum (storage space) zum Bereithalten von Informationen durch Memorieren und
 - (b) einen Arbeitsraum (operating space), der eine zentrale Informationsverarbeitungseinheit zur Speicherung und Transformation der Daten darstellt

Nach diesem Modell entwickelt sich die strukturelle Kapazität nach dem zweiten Lebensjahr nicht mehr weiter, vielmehr erfolge eine Zunahme der funktionellen Kapazität.

- (3) *KZS als aktiver Teil des LZS (nach DEMPSTER)*: dabei werden drei Dimensionen unterschieden:
 - (a) die Aktivationsenergie, die als maximale Anzahl gleichzeitig aktivierbarer Repräsentationen definiert wird (mehrere unabhängige, modalitätsspezifische Kapazitäten)
 - (b) die Verarbeitungs- bzw. Aktivationsenergie als Indikator für die Automatisiertheit von Informationsverarbeitungsprozessen
 - (c) die Vergessens- bzw. Deaktivationsrate

Veränderungen in der Verarbeitungskapazität sind dabei auch erfahrungsbedingt. Die Verarbeitungskapazität stellt innerhalb eines Entwicklungsprozesses einen Filter für gedächtnisrelevante Erfahrungen dar (vgl. KAIL, 1992).

- (4) SCHNEIDER & PRESSLEY (1997) weisen in einer Übersicht vor allem darauf hin, dass die Geschwindigkeit der Itemidentifizierung (Informationsverarbeitung) und die Automatisierung von Itemsequenzierungen wichtige Determinanten der Gedächtniskapazität darstellen und dass diesbezügliche entwicklungsbedingte Rückstände wenigstens teilweise die damit verbundenen Unterschiede in der Gedächtnisspanne erklären, was eine effektivere Nutzung von Operationen des KZS ermöglicht.

(5) Nach Ansicht von HULME & MacKENZIE (1992) ist der Zuwachs der Gedächtnisspanne mit dem Alter als Hinweis einer allgemeinen Zunahme der Fähigkeit zu betrachten, Informationen aufzunehmen und diese zu verarbeiten. Diese Zunahme der Gedächtnisspanne ist demnach nicht in einer Veränderung der Speicherkapazität per se zu sehen, sondern gründet sich vielmehr in einer verbesserten Verarbeitungseffizienz.

Die Entwicklung struktureller und funktioneller Kapazitäten lernbehinderter Kinder und Jugendlicher im Rahmen von Qualität, Struktur und anderen Erinnerungsleistungen sowie Zeugenaussagen ist wenig erforscht und stellt somit ein Ziel der vorliegenden Arbeit dar.

Die am Erinnerungsprozess beteiligten Phasen sind zu unterteilen in (1) Enkodierung, (2) Speicherung und (3) Abruf von Informationen, die nun jeweils kurz erläutert werden sollen:

Als *Enkodierung* wird jener Prozess verstanden, bei dem das Nervensystem ein reales Geschehen im Gedächtnissystem entsprechend repräsentiert (vgl. MILNE & BULL, 2003). Bei Betrachtung eines physikalischen Objekts oder eines Ereignisses wird diese Information von den sensorischen Rezeptoren im Auge aufgenommen und in Form von neuronalen Erregungen enkodiert. Neurophysiologische (elektrische) Signale werden dann dem Gehirn übermittelt, um dort im Folgenden weiter verarbeitet werden zu können. Die so entstehenden Repräsentationen im Gehirn können von phonetischer Art sein oder sie werden in Form von Wörtern oder mentalen Bildern kodiert.

Der Enkodierung folgt im nächsten Schritt die *Speicherung* der kodierten Information, die nun ins Gedächtnissystem integriert und dort abgelegt wird. Die Form der Speicherung hängt dabei von der Art des Materials ab, das erinnert werden soll, sowie von den besonderen Techniken, die bei der Enkodierung angewendet wurden.

Der *Abruf* der enkodierten und gespeicherten Information stellt die letzte Phase des Erinnerungsprozesses dar. Die Information muss nun aus dem Gedächtnis ins aktuelle Bewusstsein zurückgeholt werden. Dabei kann es zu Problemen kommen, wenn Informationen zwar im Gedächtnis gespeichert wurden, aber nicht abrufbar sind (vgl. MILNE & BULL, 2003).

Nach der dualen Kodierungstheorie von PAIVIO sollen zur Interpretation episodischer Behaltens- und Erinnerungssituationen zwei voneinander unabhängige, aber teilweise miteinander verbundene Kodierungssysteme an der Enkodierung, Speicherung, Organisation sowie am Abruf von Informationen beteiligt sein. Das imaginale System ist dabei für die Verarbeitung und Repräsentation perzeptueller Informationen zuständig sein, ein verbales System dagegen für linguistische Informationen mit der Möglichkeit zur Kommunikation. Auf verschiedenen Ebenen der Informationsverarbeitung unterscheidet Paivio ein repräsentationales, ein referentielles und ein assoziatives Verarbeitungsstadium (vgl. WIPPICH, 1984, Bd.1). WIP-

PICH (1984) verweist auf die Klasse der so genannten *dualen Prozess-Theorien*, wonach Reproduktionen zwei interne Prozesse implizieren. In einer ersten Stufe, dem Generierungsprozess (auch Such- bzw. Abrufphase) werden mögliche Antwortalternativen gesucht und generiert, in einer zweiten Stufe sollen die potentiellen Antworten überprüft werden (interner Entscheidungs- bzw. Diskriminationsprozess). Nach heutigem Kenntnisstand müssen PAIVIO's Ausführungen ergänzend zu einer *multiplen Kodierungstheorie* führen, in der Informationen zum einen auf verschiedenen Ebenen repräsentiert sein können und in der zum anderen neben sprachfreien, abstrakten Repräsentationen auch bildhafte, motorische, emotionale, motivationale und andere Bedeutungskomponenten relevant sind.

Demgegenüber besagt das von TULVING (1973) formulierte Prinzip der *Enkodierungsspezifität*, dass episodische Informationen ereignisspezifisch sind und reaktualisiert werden können, wenn Abrufinformationen eine Nutzung der episodischen Spuren ermöglichen. Demzufolge resultieren Behaltensleistungen aus dem Zusammenspiel und Zusammenpassen zwischen episodischen Spuren einerseits und Abrufinformationen andererseits. Enkodierungsspezifität ist generell auf Kontexteffekte angelegt, d.h. jede Enkodierung ist kontextspezifisch und damit jedoch bei Erinnerungsversuchen auch kontextanfällig. Dabei kann nach BADDELEY zwischen so genannten interaktiven Kontexten (Einfluss der Kontextinformationen auf die Interpretation und Enkodierung) und unabhängigen Kontexten unterschieden werden (vgl. WIPPICH, 1984, Bd.2).

Sollen nun Informationen oder Geschehnisse erinnert werden, kann es zu verschiedenen Arten von Fehlern kommen. Diese können nicht nur beim Abruf von Informationen auftreten, sondern auch in den anderen beiden Prozessen der Erinnerung. Wenn man sich an Informationen nicht mehr erinnern kann, könnte es z.B. daran liegen, dass sie ursprünglich gar nicht wahrgenommen wurden, das Problem also in mangelnder Aufmerksamkeit begründet liegt. Möglicherweise wurde die Information zwar wahrgenommen, jedoch nicht kodiert oder sie wurde zwar kodiert, aber falsch oder gar nicht gespeichert (vgl. MILNE & BULL, 2003).

Zur Bestimmung und Vorhersage kognitiver Fähigkeiten eines Menschen sind die Stufe(n) der Enkodierung, eine voraufmerksame Verarbeitung sowie die Kapazität des Arbeitsgedächtnisses von Bedeutung. Das so genannte '*preattentive processing*', das als unwillkürlich, unbewusst und automatisch beschreiben wird, führt nach SCHWEIZER (1995) zwar zur einer Beschleunigung der Reaktionszeit, trägt jedoch nicht direkt zur kognitiven Fähigkeit bei, wohingegen die anderen beiden genannten elementaren Prozesse diesbezüglich besonders relevant sind. Bei wichtigen Einzelheiten und visueller Informationsverarbeitung sind umfassende Enkodierungsprozesse erforderlich, die im letzteren Fall auch figurale Informationen bereitstellen (vgl. SCHWEIZER, 1995).

Neben den oben beschriebenen Aspekten des Speichermodells des Gedächtnisses sei neben dem assoziativen Gedächtnis insbesondere auf die *Rekonstruktivität des Gedächtnisses* hingewiesen, die bereits auf BARTLETT (1932) zurückgeht. Demnach müssen enkodierte und gespeicherte Spuren oder Engramme nicht nur mehr gefunden und dekodiert werden, sondern es wird jeder Gedächtnisinhalt im Augenblick des Erinnerns durch einen aktiven Konstruktionsprozess anhand von einfachen Schemata und Stützstellen neu aufgebaut. Dabei auftretende Veränderungen und Verzerrungen, die Neu- und Umbildung von Ordnung und Strukturen, Kontaminationen, Ergänzungen sowie die Flexibilität der Bedeutungszuweisung und Aufhebung von Mehrdeutigkeit sind durch viele Experimente belegt. Rekonstruktionsprozesse beziehen sich dabei nicht auf das ursprünglich erlebte Ereignis oder die ursprüngliche Beobachtung, sondern auf die jeweils am kürzesten zurückliegende Aktualisierung der Vorfälle etwa in einer Vernehmung (vgl. GREUEL et al., 1998).

1.1.2 Relevante Bereiche des Gedächtnisses für Zeugenaussagen

Durch die bisherigen Ergebnisse der Lern- und Gedächtnisforschung lassen sich die für die forensische Psychologie und für die Glaubhaftigkeitsbeurteilung relevanten Bereiche wie folgt aufgliedern (vg. GREUEL et al., 1998):

Kurzzeitgedächtnis - *Langzeitgedächtnis*
Prozedurales Gedächtnis - *Deklaratives Gedächtnis*
Semantisches Gedächtnis - *Episodisches Gedächtnis*
Intentionales Lernen - *Inzidentelles Lernen*
Implizites Gedächtnis - *Explizites Gedächtnis*
Wiedererkennen - *Reproduktion*

Da in der forensischen Praxis und Realität Zeugen ein beobachtetes Ereignis oft erst einige Stunden, oftmals sogar Wochen oder Monate nach dem Erlebnis oder Geschehen dies erinnern bzw. erstmals darüber berichten müssen, richtet sich bei Zeugenaussagen das Augenmerk vorwiegend auf das *Langzeitgedächtnis*, da sich das Kurzzeitgedächtnis nur auf die unmittelbare Wiedergabe eines beobachteten Ereignisses bezieht.

Die Unterscheidung zwischen deklarativem und prozeduralem Gedächtnis geht auf Anderson (1983) zurück. Unter *deklarativem Gedächtnis* versteht man die Fähigkeit zur Reproduktion von Wissenstatbeständen und kognitiven Operationen, während das *prozedurale Gedächtnis* für die Reproduktion von Handlungs- und Bewegungsabläufen relevant ist. Für die

Aussage eines Zeugen ist meist nur deklaratives Wissen von Bedeutung (vgl. GREUEL et al., 1998).

Die Differenzierung zwischen semantischem und episodischem Gedächtnis geht bereits auf Tulving (1972) zurück. Dabei werden im *episodischen Gedächtnis* spezifische, persönliche Ereignisse und Erfahrungen gespeichert, während sich die Inhalte des *semantischen Gedächtnisses* auf das so genannte Faktenwissen beziehen, also auf bestimmte bedeutungsvolle Sachverhalte, unabhängig von der konkreten Bedingung unter der sie erlebt wurden. Da nach den Grundsätzen des deutschen Strafverfahrens Zeugen nur diejenigen Personen sind, die über eigene unmittelbare Erlebnisse berichten, sind für Zeugenaussagen lediglich episodische Gedächtnisinhalte relevant, die an spezifische autobiographische Ereignisse gekoppelt sind (vgl. GREUEL et al., 1998).

Oftmals rechnen Zeugen, die ein Ereignis beobachten, in diesem Moment nicht damit, diese Beobachtung später einmal vor Gericht schildern zu müssen. Sie beobachten daher eher beiläufig, ohne wirklich aufmerksam auf bestimmte Details zu achten. Zeugenaussagen geht daher in den meisten Fällen kein intentionales, sondern ein *inzidentelles Lernen* voraus (vgl. GREUEL et al., 1998).

Weiterhin können das implizite und das explizite Gedächtnis unterschieden werden. Unter *explizitem Gedächtnis* werden Gedächtnisinhalte verstanden, derer sich eine Person im Gegensatz zum impliziten Gedächtnis bewusst wird und die bei Abfrage gezielt repliziert werden können (vgl. GREUEL et al., 1998).

In der Aussagepsychologie sind in erster Linie die freie *Reproduktion* (free recall, freier Bericht) aber auch das *Wiedererkennen* von Bedeutung. In der freien Reproduktion eines Sachverhaltes werden hohe kognitive Anforderungen an die betroffene Versuchsperson gestellt, da sie freie, nach eigenen Strukturierungsregeln gestaltete, Informationen liefern muss, ohne vom Interviewer inhaltliche Vorgaben zu erhalten. Bei der Anwendung der Methode des Wiedererkennens ist die Gedächtnisleistung in der Regel deutlich besser als bei der Anwendung der Methode des freien Berichts. Ein Zeuge muss ein beobachtetes Ereignis entweder aktiv im Gedächtnis reproduzieren oder er muss Beschuldigte oder tatrelevante Beweisstücke identifizieren (vgl. GREUEL et al. 1998).

Für Zeugenaussagen sind vorrangig jene Aspekte von Bedeutung, die ganz allgemein dem Langzeitgedächtnis zuzuordnen sind, deklaratives Wissen repräsentieren, primär inzidentell gelernt wurden, die als explizite Gedächtnisinhalte bewusst gemacht werden können, sich auf konkrete Episoden aus der eigenen Biographie beziehen und frei bzw. aktiv reproduziert werden müssen (vgl. GREUEL, 2001).

In diesem Zusammenhang stellt sich auch die Frage bzw. ist von Bedeutung, welche Gedächtnisprozesse unterschiedlichen Aussagetypen zugrunde liegen (vgl. GREUEL, 2001):

Tabelle 1-1: Gedächtnisprozesse und Aussagetypen

Gedächtnisprozess	Aussagetypus
autobiographisches Gedächtnis autonoetisches Gedächtnis Handlungsgedächtnis	erlebnisbasierte Aussage
Semantisches Gedächtnis Textgedächtnis	Phantasie-Aussage
Textgedächtnis Konversationsgedächtnis	induzierte Aussage
noetisches Gedächtnis	beobachtungsbasierte Aussage (Übertragungen)

1.1.3 Alltagsgedächtnis

Neben der traditionellen laborexperimentellen Grundlagenforschung hat sich in den letzten Jahren ein weiterer, für die forensische Praxis bedeutsamer Forschungszweig entwickelt, der sich mit den Besonderheiten des menschlichen *Alltagsgedächtnisses* (*everyday memory*) und dem Einbeziehen natürlicher Gedächtnisprozesse in die wissenschaftliche Theoriebildung befasst. Demnach sollte die bisherig dominierende Forschung bezüglich des Settings der laborexperimentellen Betrachtung verbaler Lernprozesse zwingend um Aspekte des Kontextes, individueller Bedeutsamkeiten oder interindividueller Differenzen erweitert werden (vgl. GREUEL, 2001), beispielsweise im Zusammenhang mit forensisch bedeutsamen Themen oder speziellen Personengruppen. Dies soll ein Anliegen der vorliegenden Arbeit sein. Inwieweit die erhobenen Befunde wiederum eventuelle Rückschlüsse oder Auswirkungen auf die Grundlagenforschung haben, bleibt zu diskutieren. Zwei Schwerpunkte innerhalb der *‘everyday memory research’* haben sich bisher herauskristallisiert und etabliert: zum einen die Forschung zum *autobiographischen Gedächtnis*, zum anderen zum so genannten *Augenzeugengedächtnis* (*eyewitness memory*; *eyewitness testimony*). Beide Schwerpunkte bilden unter anderem und zum Teil die theoretische Grundlage dieser Arbeit.

1.1.3.1 Autobiographisches und Augenzeugengedächtnis

Bei der Befragung von Zeugen ist nach GREUEL et al. (1998) oftmals das *autobiographische Gedächtnis* von Bedeutung, da es bei der Beurteilung inkriminierter Geschehnisse im Rah-

men der Glaubhaftigkeitsbegutachtung auch auf die Einordnung in den individuellen Lebenskontext des Zeugen ankommt. Entsprechende Studien fokussieren primär auf die verschiedenen Typen bzw. qualitativen Aspekte autobiographischer Erinnerungen; sie beziehen sich auf private, persönlich bedeutsame Erlebnisse aus der individuellen Vergangenheit, die nur schwer objektiv überprüfbar sind und somit kaum externe Validierungsmöglichkeiten bieten. Beim autobiographischen Gedächtnis gilt das so genannte Pollyanna-Prinzip, demzufolge angenehme Ereignisse besser erinnert werden als unangenehme, insbesondere wenn diese weiter zurückliegen, da letztere oft mit Angst verbunden sind und Angst eine Interferenz mit der Fähigkeit des Erinnerns zeigt. An angenehme Ereignisse wird demnach häufiger gedacht, so dass ein gewisser Wiederholungseffekt wahrscheinlich ist bzw. auftritt. Weiterhin werden in diesem Zusammenhang das so genannte *Blitzlichtgedächtnis* (*'flashbulb-memory'*) und ein so genanntes Tunnelgedächtnis angeführt, demzufolge gerade schockierende Ereignisse und Katastrophen höchst detailliert und lebhaft erinnert werden. Bei letzterem sind allerdings die Kontextereignisse (kurz vor und einige Zeit nach der Katastrophe) sehr schlecht reproduzierbar. Ein ebenfalls bekanntes und in polizeilichen Ermittlungsverfahren und Gerichtsprozessen oft beobachtetes Phänomen ist das des *'Waffenfokus'* (*weapon focus*), wonach sich Zeugen und Opfer kaum bzw. sehr schlecht an den Täter erinnern konnten und dementsprechend unzureichende Personenbeschreibungen abgeben, da sich ihre Aufmerksamkeit in der jeweiligen Situation des Überfalls oder sonstigen ähnlichen Geschehnissen tunnelartig auf die Waffe richtet, von der die unmittelbare Bedrohung ausgeht (vgl. GREUEL et al., 1998). Inwieweit sich das eben beschriebene Phänomen auch auf andere Objekte in bestimmten Settings (z. B. Autos bei Verkehrsunfällen) oder spezielle Personen (z. B. Menschen mit intellektuellen Beeinträchtigungen) zumindest in experimentellen Bedingungen übertragen lässt und ob bzw. welche motivationalen Aspekte oder andere Aspekte (selektiver) Aufmerksamkeit hierbei eine Rolle spielen können, bleibt ebenfalls zu diskutieren.

Nach aktuellem Forschungsstand und derzeitiger Auffassung ist von einer hierarchischen Organisation des autobiographischen Gedächtnisses auszugehen. Auf der untersten Ebene der Gedächtnishierarchie finden sich die Repräsentationen spezifischer Erlebniserinnerungen mit einer Fülle sensorischer Reizinformationen als ereignisspezifisches Wissen wieder, die raum-zeitlich konkret verankert sind. Solche ereignisspezifischen Erinnerungen und Episoden sind sehr fragil, da sie nach einer gewissen Zeit in übergeordnete „generalisierte Erinnerungen“ integriert werden. So ist demnach sowohl von einer zeitlichen als auch thematischen Organisation des autobiographischen Gedächtnissystems auszugehen und die Erinnerbarkeit und Darstellung einzelner Episoden in Abhängigkeit von zeitlichen und/oder thematischen Organisationsprinzipien – sowohl in der Enkodierungs- als auch in der Abrufphase- zu betrachten (vgl. GREUEL, 2001). Verschiedene Untersuchungen ergaben dabei,

dass sowohl autobiographische Ereignisse als auch Themen häufig aus Episoden bestehen, die aus mehr als einem Tag entnommen sind (BURT et al., 2003).

Inhalte des autobiographischen Gedächtnisses sind um zeitliche Bezugspunkte (*temporal landmarks*) organisiert, denen eine zweifache Rolle zukommt: Zum einen handelt es sich um inhaltliche Aspekte, zum anderen haben sie strukturierende Funktion, indem sie andere Erinnerungen organisieren und effizient Hinweisreize für das Abrufen von autobiographischen Erinnerungen darstellen (vgl. GREUEL, 2001).

Nach WELCH-ROSS (1995) beginnen Kinder während des Vorschulalters differenzierte Kontextinformationen im autobiographischen Gedächtnis zu repräsentieren und den Nutzen von Erinnerungen für eine Quelleninformation zu verstehen. Für die Entwicklung des autobiographischen Gedächtnisses ist daher von besonderer Bedeutung unterscheiden zu können, ob Erinnerungen und Gedächtnisinhalte auf dem persönlichen Erfahren eines Ereignisses beruhen oder ob diese nur aus Erzählungen oder der Vorstellung eines Ereignisses stammen (vgl. WELCH-ROSS, 1995). Zur Verbesserung des Systems des autobiographischen Gedächtnisses sind zum einen ein Bewusstsein über die verschiedenen Möglichkeiten wie Wissen erworben wird erforderlich, zum anderen ein Verständnis von fiktionalen mentalen Repräsentationen. Diese beiden Fähigkeiten können auch zur kognitiven Untermauerung des Verständnisses von Quellenidentifikation beitragen. Daneben sind noch andere Prozesse von Bedeutung wie etwa die Struktur der elterlichen Konversation hinsichtlich Erinnerungen (elaborativ oder repetitiv) oder die Art von Inhalten auf denen der Focus liegt (etwa spezifisch emotionale Inhalte).

Als maßgebliche Komponente des autobiographischen Gedächtnisses gilt die *visuelle Repräsentation*. Das Vorhandensein von mentalen Bildern (*mental images*) liefert Menschen letztendlich auch den primären phänomenologischen Hinweis, um zwischen selbst erlebten und lediglich noetischen Erinnerungen differenzieren zu können – ein Aspekt, der ebenfalls im Modell des 'reality monitoring' von Bedeutung ist und noch dargestellt wird (vgl. GREUEL, 2001).

Weiterhin stellt sich autobiographisches Erinnern im Sinne des *partiellen Konstruktivismus* als Kombination von Speicherung und Rekonstruktion dar, wonach zentrale Elemente einer subjektiv bedeutsamen Erfahrung in einem perzeptuellen Bildformat gespeichert und über längere Zeitintervalle relativ genau behalten werden. Eher periphere Details unterliegen dahingegen vermehrt rekonstruktiven Veränderungsprozessen (vgl. GREUEL, 2001).

Nach GREUEL (2001) haben Studien zum autobiographischen Gedächtnis vor allem solche Erinnerungen zum Inhalt, die für das subjektive Selbst bedeutsame, identitätsstiftende, persönliche Erfahrungen darstellen (Reproduktion von aus der Akteur – Perspektive gemachten Erfahrungen). Zudem seien Tunnel-Erinnerungen, die zwar die Gedächtnisleistung für zentrale Details erhöhen, begrenzt auf emotional-negative Ereignisse

für zentrale Details erhöhen, begrenzt auf emotional-negative Ereignisse (vgl. auch BERNTSEN, 2002; BERNTSEN, 2001; NIEDZWIENSKA, 2003).

D'ARGEMBEAU et al. (2003) untersuchten Erinnerungsqualitäten von positiven, negativen und neutralen autobiographischen Ereignissen und kamen zum Ergebnis, dass positive Erinnerungen mehr sensorische und kontextuelle Einzelheiten aufwiesen als negative oder neutrale, wobei sich negative und neutrale Erinnerungen in den meisten Dimensionen nicht besonders unterschieden. Zum einen werde Emotionales besser, d.h. facettenreicher enkodiert als neutrale Geschehnisse, andererseits ist hierfür auch ein höheres Maß an Aufmerksamkeit erforderlich. Jedoch werden diese Prozesse und Informationen diesbezüglich quasi freiwillig eher bewusst und sorgfältiger elaboriert, da emotional bedeutsame Ereignisse von größerem persönlichen Interesse sind als neutrale. Zudem wird an emotionale Ereignisse öfter gedacht und diese werden meistens auch öfter anderen Personen mitgeteilt und profitieren demnach von spezifischen neuronalen Konsolidierungsprozessen. Außerdem fanden die Autoren, dass emotionale und neutrale Ereignisse aus unterschiedlichen Perspektiven heraus erinnert werden und Personen, die Gefühle eher unterdrücken, sich mit Ausnahme von qualitativen Merkmalen ihrer negativen Erinnerungen diesbezüglich nicht von anderen Menschen unterscheiden (vgl. D'ARGEMBEAU et al., 2003).

Der Focus der Untersuchung von BLUCK & LI (2001) lag auf der Frage, ob unterschiedliche Faktoren die anfängliche bzw. die finale Erinnerungsleistung vorhersagen könnten. Dabei zeigte sich, dass für die anfängliche Erinnerungsleistung vor allem negative emotionale Reaktionen während des Ereignisses gute Prädiktoren sind, dagegen für die finale Erinnerungsleistung vor allem die Häufigkeit des 'einem Ereignis ausgesetzt Seins'.

Im autobiographischen Gedächtnis besteht außerdem eine Tendenz, Details von mehreren ähnlichen Ereignissen zu einem gut erinnerbaren Ereignis zu kontaminieren. Diesen Effekt des Verlustes ereignisspezifischer Details zugunsten einer Schematisierung von erinnerten Ereignissen hat NEISSER (1981) als *repisodisches Gedächtnis* bezeichnet. So besteht nach GREUEL et al. (1998) die Möglichkeit, dass Details von Ereignissen, die häufig vorgekommen sind, aufgrund einer Schematisierung den jeweiligen einzelnen Ereignissen nicht mehr zuverlässig zugeordnet werden.

Dass die gerade ausgeführten Befunde auch auf das Augenzeugengedächtnis übertragbar sind, ist wahrscheinlich.

Das *Augenzeugengedächtnis* befasst sich im Hinblick auf den forensischen Verwertungszusammenhang fast ausschließlich mit Fragen der Vollständigkeit und Genauigkeit von Zeugenaussagen. Im Zentrum der Forschung zum Augenzeugengedächtnis stehen öffentliche bzw. im Forschungssetting inszenierte Beobachtungsvorfälle, die entsprechend extern validiert werden können. Damit ergeben sich auch unterschiedliche Erinnerungsintervalle für

die in den jeweiligen Forschungskontexten erhobenen Gedächtnisreproduktionen. Während das Erinnerungsintervall bei autobiographischen Erinnerung lang ist, ist es bei den Beobachtungen des Augenzeugengedächtnis meist kurz bis mittel. Zudem beschäftigt sich die Forschung zum Augenzeugengedächtnis überwiegend mit Berichten über persönliche Beobachtungen, denen gemeinhin der eigene Handlungsbezug fehlt (gedächtnismäßige Repräsentationen von Beobachtungsleistungen) (vgl. GREUEL, 2001).

In ihrer Untersuchung überprüften NAVEH-BENJAMIN et al. (2003) die so genannte *‘assoziative Defizit-Hypothese’ (ADH)*. Demnach kann der überwiegende Teil des Defizits von älteren Menschen bezüglich der episodischen Erinnerungsleistung auf Schwierigkeiten zurückgeführt werden, Attribute und Informationseinheiten ohne Bezug in eine zusammenhängende Einheit zu integrieren. Dies gilt vor allem auch für Bilder als Stimulusmaterial. Demgegenüber zeigte eine Kontrollgruppe junger Erwachsener, die die Informationen unter Bedingungen geteilter Aufmerksamkeit enkodierten, dieses assoziative Defizit nicht (vgl. NAVEH-BENJAMIN et al., 2003).

1.1.3.2 Der Begriff des Augenzeugen

Nach HÖFER & KÖHNKEN (1998) sind *Zeugen* Menschen, die Vorfälle, Ereignisse, Personen und Handlungen beobachten, die von juristischer Relevanz sind. Grundsätzlich, sowie im Rahmen der forensischen Aussagepsychologie, die sich speziell mit der Erforschung des Realitätsgehalts von Zeugenaussagen beschäftigt, gilt als wesentliche Hauptfunktion des Zeugen, Ereignisse aus einer retrospektiven Perspektive zu schildern.

Trotz solcher Definitionen ist der Begriff des *‘Augenzeugen’* im Hinblick und auf dem Hintergrund noch ausstehender Theorienbildung zum so genannten Augenzeugengedächtnis vage, mehrdeutig und vielschichtig und bedarf nach GREUEL (2001) einer weiteren terminologischen und konzeptuellen Differenzierung.

GREUEL (2001) unterscheidet demnach drei unterschiedliche Typen von Zeugen:

Zum einen werden *unbeteiligte Augenzeugen* von beteiligten Augenzeugen begrifflich so abgegrenzt, dass ein unbeteiligter Zeuge derjenige ist, der rechtlich relevante Interaktionen beobachtet und diese persönlichen Wahrnehmungen im forensischen Kontext frei reproduziert. Bestimmte Personen und/oder Objekte zu beschreiben oder sie zu identifizieren, können weitere Leistungen eines unbeteiligten Augenzeugen darstellen. Demgegenüber gilt als *beteiligter Augenzeuge* derjenige, der als Opfer unmittelbar und aktiv an einem fraglichen Geschehen involviert war und als so genannter *‘Opfer – Zeuge’* über diese persönlichen

Erlebnisse aussagt oder aber spezifische Beschreibungen oder Identifikationen vornimmt. Somit wird über Sachverhalte berichtet, die subjektiven Erlebniswirklichkeiten aus der individuellen Vergangenheit des Zeugen entsprechen.

Zuletzt sollte darüber hinaus der bis dato von der rechtspsychologischen Forschung wenig beachtete, jedoch gerade besonders relevante Typus des 'Täter – Zeugen' zunehmend von Interesse sein, „der insofern einen Sonderstatus einnimmt, als dass er als (Mit-) Täter eines Verbrechens Aussagen zum fraglichen Tatablauf macht und dabei sehr spezifischen motivationalen Bedingungen unterliegt (z.B. Kronzeugen)“ (GREUEL, 2001, S. 63). Nachstehende Tabelle fasst die verschiedenen Zeugentypen und ihre Aussageleistungen zusammen.

Tabelle 1-2: Differenzierung nach Zeugengruppen und Aussageleistungen

Zeugengruppe	Aussagegegenstand	Aussageleistung
unbeteiligter Augenzeuge („bystander“)	Aussagen über persönlich Wahrgenommenes	Reproduktion von Interaktio- nen/Ereignissen, Beschreibung von Personen oder Objekten, Identifizierung durch Wiedererkennen
beteiligter Augenzeuge („Opferzeuge“)	Aussagen über persönlich Erlebtes	Reproduktion von Interaktio- nen/Ereignissen, Beschreibung von Personen oder Objekten, Identifizierung durch Wiedererkennen
beteiligter „Täter – Zeuge“	Aussagen über persönlich Erlebtes	Reproduktion von Interaktio- nen/Ereignissen, Beschreibung von Personen oder Objekten, Identifizierung durch Wiedererkennen

(aus GREUEL, 2001, S. 64)

1.1.4 Generelle Einflussfaktoren auf die Gedächtnisleistung

Zahlreiche unterschiedliche Faktoren können auf die Gedächtnisleistung von Zeugen einwirken und auf das Erinnern komplexer Ereignisse Einfluss nehmen. Dabei können diese Faktoren verschiedenen spezifischen Phasen im Gedächtnis- oder Erinnerungsprozess (Enkodierung, Speicherung und Abruf) zugeordnet werden, was jedoch nicht immer zwingend und eindeutig der Fall sein muss.

Im Zusammenhang mit der *Enkodierungsphase* sind z.B. *äußere Faktoren*, wie Beobachtungsdauer, Entfernung vom Tatort oder Lichtverhältnisse zu nennen, die sich direkt auf die Erinnerung an das Ereignis auswirken können, da die Qualität einer Erinnerung von der Interaktion zwischen dem Augenzeugen und dem zu erinnernden Ereignis abhängt (vgl. KELLOWAY et al., 2004; DAVIES, 1993). Weiterhin stellt auch der *allgemeine Zustand des Zeugen* einen Faktor dar, der die Qualität von Erinnerungen beeinflussen kann, d.h. ob der entsprechende Zeuge bei der Beobachtung eines Ereignisses unter dem Einfluss von Alkohol oder anderen psychoaktiven Substanzen steht oder stand. *Stress* ist ein weiterer Faktor, der auf die Gedächtnisleistung von Zeugen einen Einfluss haben kann. Aufgrund der damit verbundenen ethischen Fragen ist es nicht ganz einfach, den Zusammenhang zwischen Stress und Gedächtnisleistung zu erforschen. In den bisher durchgeführten Studien konnte zusammenfassend ein komplexes Wechselverhältnis nachgewiesen werden, das von einer Reihe von Faktoren beeinflusst werden kann. Die Forschungsergebnisse gestalten sich allerdings unterschiedlich. Einige Studien kommen zu dem Schluss, der Zusammenhang zwischen Stress und Genauigkeit bei der Wiedergabe verlaufe umgekehrt U – förmig, wobei die höchste Genauigkeit bei einem mittleren Stress – Niveau erreicht wird (Deffenbacher, 1983); verschiedene andere Studien konnte nachweisen, dass Stress die Menge an erinnerter Information (Quantität) erhöht und die Erinnerungsgenauigkeit (Qualität) verbessern kann (vgl. z.B. CHRISTIANSON & HUBIETTE, 1993). Ein Grund dafür könnte sein, dass Stress ein Faktor ist, der ein Ereignis einzigartig macht und man es sich dadurch umso besser ins Gedächtnis einprägen kann (YUILLE & TOLLESTRUP, 1992). Daneben kann die *Aufmerksamkeit*, die ein Zeuge auf bestimmte Einzelheiten eines Ereignisses richtet, ebenfalls seine Erinnerungsleistung beeinflussen. Hier sei an das schon erwähnte Phänomen des *‘weapon focus’* erinnert (vgl. GREUEL et al., 1998; MILNE & BULL, 2003). Bezüglich der (*emotionalen*) *Betroffenheit* von Zeugen findet sich in der wissenschaftlichen Literatur ebenfalls keine eindeutige Befundlage. So sei die Erinnerungsleistung von Zeugen, die selbst in den Tathergang verwickelt waren, laut einiger Studien besser als die von unbeteiligten Beobachtern (vgl. COHEN & FAULKNER, 1988), allerdings sprechen auf der anderen Seite auch einige Studien dafür, dass die Betroffenheit von Zeugen keine Auswirkungen auf deren Gedächtnisleistungen hat (vgl. z.B. ROBERTS & BLADES, 1998). Auch das *Wissen* eines Zeugen kann seine Gedächtnisleistung für ein bestimmtes Ereignis beeinflussen, weshalb es unterschiedlichen Altersklassen (qualitativ und quantitativ) andere Gedächtnisleistungen feststellbar sind: im Vergleich zu Erwachsenen verfügen Kinder und Jugendliche noch weniger über bestimmtes Vorwissen und können sich somit auch an weniger Details erinnern bzw. begehen in der Konsequenz auch andere Arten von Fehlern (vgl. auch DAVIES, 1993). Weiterhin finden sich einige Belege dafür, dass nicht die reine Anzahl von verbalen Wiederholungen, sondern der Zeitpunkt der Wiederholungen von entscheidender Bedeutung für das mittel-

und langfristige Behalten von (bedeutsamen) Erfahrungen ist. Je schneller die erste verbale Wiederholung und damit auch Befragung erfolgt, desto besser ist die langfristige Behaltensleistung (Fivush & Schwarzmüller, 1998; Poole & White, 1995; Shrimpton et al., 1998; Warren & Lane, 1995; zit. nach GREUEL, 2001). Demnach führen die Verbalisation und Wiederholung von Erlebniserinnerungen in erster Linie in der Enkodierungsphase zu einer Verbesserung der Behaltensleistung.

Was den *Prozess der Speicherung von Informationen* angeht, ist insgesamt festzustellen, dass sich die Gedächtnisleistung systematisch mit zunehmender zeitlicher Verzögerung verschlechtert. Allerdings sind diese Veränderungen negativ beschleunigt, das Ausmaß der Veränderung wird also mit zunehmender Verzögerung immer kleiner. Demnach wäre zu vermuten, dass Zeugenaussagen fast wertlos sind, da bereits nach kürzester Zeit mehr als die Hälfte der eingprägten Erinnerungen verloren gegangen sind. Dabei muss jedoch berücksichtigt werden, dass in den entsprechenden Studien, auf die sich diese Befunde beziehen, häufig Silben als Versuchsmaterial verwendet wurden und dieses Vorgehen auch anfällig dafür ist, proaktive Interferenz zu erzeugen (vgl. DAVIES, 1993). Dagegen beziehen sich Zeugenaussagen jedoch in der Regel auf sinnvoll erlebte Ereignisketten oder autobiographische Erlebnisinhalte (vgl. GREUEL, 1998).

Nach MILNE & BULL (2003) kann auch der *Abruf* von gespeicherten Informationen bestimmten Einflussfaktoren unterliegen. So ist beispielsweise der *konstruktive Charakter der Gedächtnisses* dafür verantwortlich, dass ein komplexes Ereignis, das erinnert werden soll, erst mithilfe der kodierten Aspekte sowie der über die Welt im Allgemeinen vorhandenen Informationen rekonstruiert werden muss. Je nachdem welche allgemeinen Informationen beim Abruf zur Verfügung stehen, wird die Aussage der Person über ein vergangenes Ereignis ausfallen. Anhand einer klassischen Untersuchung konnten Loftus und Palmer (1974) nachweisen, dass auch *voreilige Schlussfolgerungen* dazu führen können, dass beim Abruf Dinge erinnert werden, die gar nicht beobachtet wurden. So führte in der genannten Studie die Veränderung von nur einem Wort in einer Frage dazu, dass Versuchspersonen sich an Details erinnern, die in einer vorangegangenen präsentierten Diaserie über einen Autounfall gar nicht dargeboten wurden. Zudem konnte in einigen Studien nachgewiesen werden, dass Erinnerungen an Ereignisse mit persönlicher Bedeutung für die jeweiligen Opfer durch Rationalisierungen im Nachhinein verzerrt werden, also in manchen Situationen oder entsprechenden Zeugenaussagen oder Schilderung von Geschehnissen von einer gewissen *Parteilichkeit* ausgegangen werden kann und muss (vgl. BOON & DAVIES, 1996).

Neben emotionalen und motivationalen Faktoren - sei es zum Zeitpunkt der Beobachtung oder aber zum Zeitpunkt der Abrufsituation, die sich positiv wie negativ auf diesen auswirken kann - sei abschließend der *Kontext*, in dem die Erinnerung kodiert wurde, zu nennen, der nach einhelliger aktueller Meinung als positiv für den Abruf von Informationen anzusehen ist.

Befindet man sich am gleichen Ort oder im gleichen Kontext wie zu dem Zeitpunkt, in dem die Kodierung ursprünglich stattgefunden hat, wird die Gedächtnisleistung beim Abruf verbessert.

Individuelle Differenzen in Gedächtnisleistungen sind neben Unterschieden in Enkodierungs- und Abrufstrategien vor allem in Zusammenhang mit dem Konstrukt der kognitiven Flexibilität der Informationsverarbeitung zu betrachten.

Daneben spielen verschiedene *Persönlichkeitsdimensionen* wie Introversion, Ängstlichkeit, Selbstvertrauen aber auch generell der Umgang bzw. die Reaktionen der sozialen Umwelt eine wichtige Rolle. So findet sich die These, ängstliche Menschen tendierten dazu, Informationen eher oberflächlich zu verarbeiten. Gemäß der '*semantischen Defizithypothese*' werde die Aufmerksamkeit bei höherer Erregung von semantischen Merkmalen zu bearbeitender Information abgelenkt und somit eine tiefere Enkodierung behindert, ähnlich dem Ansatz des 'levels-of-processing' (vgl. WIPPICH, 1984, Bd.2).

In diesem Zusammenhang und speziell auch von Interesse in der vorliegenden Arbeit finden auch immer mehr der Aspekt der '*schemageleiteten Erinnerung*' sowie die Begriffe '*Schema*' und '*Skript*' selbst zunehmende Verbreitung. Nach GREUEL et al. (1998) geht das Schema-Konzept davon aus, dass wir im Laufe unseres Lebens immer mehr abstrakte Vorstellungen über verschiedene Sachverhaltsbereiche herausbilden. Schemata, die sich auf Ereignisse beziehen nennt man Skripte (Ereignis-Schemata), wobei die jeweiligen Schemata bereits die Wahrnehmung beeinflussen und später das Erinnern. Skripte sollen unter anderem das Verstehen von 'event-based situations' und die Interpretation alltäglicher Situationen erleichtern und können – wie andere kognitive Strukturen, Schemata oder Konzepte – bei der Verarbeitung 'neu einlaufender' Information aktiviert werden, sofern zwischen dem neuen Informationsangebot und schon vorhandenem Wissen oder Vorstellungen ein hinreichendes Maß an Übereinstimmung (goodness-of-fit) vorliegt (vgl. WIPPICH, 1984, Bd.2). Zudem dienen Skripte dazu Erwartungen zu lenken, Schlussfolgerungen zu ziehen, Informationen zu verarbeiten und Gedächtnislücken zu füllen. Sich eher auf Skripte während der Erinnerung zu verlassen als auf die ursprüngliche Gedächtnisspur ist dann wahrscheinlicher, wenn Elemente aus dem Skript zentral bzw. mit dem Skript selbst eng verbunden sind und je länger das Behaltensintervall ist (GREENBERG et al., 1998). Dabei können Schemata wie Skripte sowohl positive wie negative (z. B. falsche Schlussfolgerungsprozesse) Effekte haben, Interpretationen beeinflussen und Elaborationen ermöglichen. Im späteren Verlauf, wenn Erinnerungen bereits unklar und verschwommen sind, haben Menschen somit die Möglichkeit, sich auf das Skript zu beziehen um Gedächtnis- und Erinnerungslücken zu schließen. Diese positive, für die Erinnerung hilfreiche 'Lückenfüllfunktion' stellt auf der anderen Seite jedoch auch eine Fehlerquelle vor allem in Form von *skriptbasierten Intrusionen* dar (vgl. GREENBERG et al., 1998). Man geht heute davon aus, dass schemadiskre-

pante Details kurzfristig besser und langfristig schlechter erinnert werden als schemakonsistente Details (vgl. GREUEL et al., 1998).

Schemata (als Bestandteil des semantischen Langzeitspeichers) sollen die Verarbeitung einlaufender (episodischer) Informationen beeinflussen, indem diese Informationen selektiv bearbeitet werden. Weitere Abstraktions-, Rekonstruktions-, Integrations- und Interpretationsprozesse reduzieren, steuern und organisieren den weiteren Erinnerungsprozess (vgl. WIPPICH, 1984, Bd.2; GREENBERG et al., 1998).

Schematheoretisch betrachtet verfügen ältere Kinder über elaborierte '*general event memories*' und damit über effizientere Organisationsstrukturen in der Enkodierungsphase. Demnach ist es erst auf dem Hintergrund generalisierter Erinnerungen in Form von Skripts möglich, Spezifisches wahrzunehmen und zu enkodieren (Ereignis-Schemata). Dabei werden Alterseffekte sowohl durch die Komplexität als auch die Vertrautheit (komplexe, differenzierte Ereignis-Schemata als besondere Form domänenspezifischen Wissens) des angesprochenen Schemas bzw. Skripts moderiert. Je unvertrauter und komplexer das übergeordnete Ereignisschema ist, desto schwieriger fällt die Enkodierung und genaue Rekonstruktion distinkter Episoden (vgl. GREUEL, 2001).

Nach RUBIN & BERNTSEN (2003) gilt die Vorstellung eines kulturübergreifenden '*life-scripts*' zwar für positive, jedoch nicht für negative Ereignisse, die Behaltensprozesse strukturieren (vgl. RUBIN & BERNTSEN, 2003).

Verschiedene Untersuchungen zur Personenwahrnehmung im Rahmen sozialer Kognitionsforschung ergaben, dass die Beschreibung, Attributionen sowie die Verarbeitung von Informationen über andere Personen als Wechselwirkungsprozess zwischen Charakteristika der Person selbst und bereits vorhandener allgemeiner Konzepte über Personen des Wahrnehmenden angesehen werden müssen (*Personen-Schemata*). Dabei können natürlich auch - in episodischen Untersuchungssituationen nachweisbare - widersprüchliche, sogenannte Inkongruenz-Effekte auftreten.

Dem (semantischen) Langzeitspeicher werden allerdings nicht nur Personen-Schemata zugeordnet, sondern auch schematische Informationen zu Berufen, sozialen Rollen oder Gruppen. *Soziale Stereotype* (soziale Schemata, Rollen-Schemata) gelten als eine Form interpersonellen Wissens, das vor allem durch Vereinfachung, Übergeneralisierung und Ungenauigkeit zu kennzeichnen ist. Stereotype erleichtern demnach zwar die Verarbeitung komplexer sozialer Informationen, was jedoch mit Fehlern und negativen Konsequenzen in anderen Bereichen einhergeht.

Insgesamt tragen Einstellungen, Skripte, Schemata und Stereotype möglicherweise zu Erinnerungsfehlern bei (KELLOWAY et al., 2004).

Grundsätzlich erinnern Kinder wie Erwachsene persönlich bedeutsame Ereignisse am besten. Dabei gilt das kindliche Gedächtnis als *idiosynkratisch*, so dass für Kinder gänzlich

andere Aspekte bzw. Ereignisse bedeutungsvoll und damit erinnerungswürdig sind als für Erwachsene. Die Behaltensleistung an sich hängt bei Kindern wie Erwachsenen im Wesentlichen von der Distinktheit des fraglichen Ereignisses ab, wobei hier in erster Linie das Ausmaß der persönlichen Bedeutsamkeit entscheidend ist. Somit werden Studien im natürlichen (Erlebnis-)Feld dieser Besonderheit offensichtlich eher gerecht. (vgl. GREUEL, 2001). Inwieweit dies auf Personen mit Intelligenzminderung zutrifft bzw. zu welchem Grad die Gedächtnisleistung auch von vom spezifischen Sachverhalt, besonderen Interessen oder dem Versuchsmaterial in der vorliegenden Untersuchung abhängt, bleibt zu diskutieren.

1.1.5 Das Konzept der Realitätsüberwachung (reality monitoring)

Neben den bereits dargestellten ist ein weiterer Gedächtnisprozess eng mit dem Wiedererkennen und dem Kontext von Ereignissen verbunden, nämlich die Zuordnung eines Vorgangs zu seinem jeweiligen Ursprung bzw. seiner Quelle. Dabei umfasst der Begriff 'Quelle' eine Vielzahl von Merkmalen, die zusammen die charakteristischen Beziehungen spezifizieren, unter denen ein Ereignis bzw. die entsprechende Erinnerung erworben wird (z.B. zeitlicher, räumlicher, sozialer Kontext, verschiedene an der Wahrnehmung beteiligte Sinne und Modalitäten). *Quellenidentifikation* bezeichnet somit die Differenzierung der Herkunft von Informationen. Solche Erinnerungssituationen erscheinen gerade im forensischen Kontext von großer Bedeutung.

Innerhalb des Modells der Realitätsüberwachung wird zwischen den Begriffen des „reality testing“, der Realitätstestung, und des „reality monitoring“, der Realitätskontrolle, differenziert. Unter Realitätstestung verstehen JOHNSON & RAYE (1981) dabei die Prozesse, die daran beteiligt sind, eine aktuelle Beobachtung eines Ereignisses von einer aktuellen Vorstellung zu unterscheiden. Im Rahmen der Realitätskontrolle oder -überwachung stellt sich die Frage, wie Menschen zwischen external und internal generierten Kognitionen und Erinnerungen diskriminieren. Auf der Ebene der gedächtnismäßigen Repräsentation geht es hier um die Abgrenzung zwischen Erinnerungen an Angetroffenes (Wahrnehmungen, Erfahrungen in der Wachwirklichkeit) und Erinnerungen an Vergegenwärtigtes (Vorstellungen, Gedanken, Pläne, Phantasien u.ä.), also beispielsweise die Frage, ob eine bestimmte Situation tatsächlich beobachtet oder sich nur vorgestellt wurde. Somit umfasst die Fähigkeit zur Realitätsüberwachung Wirklichkeitsabstufungen auf der Dimension 'external-internal' und korrespondiert damit mit der aussagepsychologischen Differenzierung zwischen erlebnisbasierten Aussagen einerseits und konstruierten Aussagen ohne Erlebnishintergrund andererseits.

JOHNSON & RAYE (1981) beschreiben die internalen, selbst-generierten Informationen ausführlicher, indem sie diese in drei verschiedene Kategorien einteilen:

In die erste Kategorie fallen die internalen Informationen, die sich darauf beziehen, dass man sich ein Ereignis, das wahrgenommen wurde, erneut vorstellt oder sich an etwas erinnert, das vor einiger Zeit geschehen ist. Dabei werden Informationen, die bereits nicht mehr im Bewusstsein oder im Arbeitsgedächtnis vorhanden sind, zu einem späteren Zeitpunkt reaktiviert, ohne dass der ursprüngliche, externe Reiz vorhanden ist.

Die zweite Gruppe innerhalb der internalen Informationen stellen die „gleichzeitigen Gedanken“ (contemporal thoughts) dar; diese beziehen sich auf die assoziativen Prozesse, die ein Ereignis, in dem Moment, in dem man es wahrnimmt, noch weiter ausschmücken.

Unter Phantasien fallen ungewöhnliche Informationen, die Bilder an Ereignisse hervorrufen, die nur in unserer Vorstellung stattfinden. Beispiele für diese Kategorie sind Träume oder das Schreiben eines Romans.

JOHNSON, HASHTROUDI & LINDSAY (1993) beschreiben das Konzept der Realitätskontrolle ausführlicher und erläutern, dass man genau betrachtet zwischen zwei unterschiedlichen Definitionen dieses Konzepts differenzieren kann:

Die erste Sichtweise unterscheidet zwischen Informationen, die von einer Person selbst, also intern generiert wurden und Informationen, die auf einen äußeren, externen Ursprung zurückgeführt werden können. Beispiele für Realitätskontrolle wären nach dieser Definition die Unterscheidung zwischen dem, was von einer Person laut ausgesprochen wurde (intern) und dem, was von einer Person gehört wurde (extern) bzw. die Unterscheidung zwischen dem, was eine Person dachte (intern) und dem, was von einer Person gehört wurde (extern). Ob es sich um eine externe oder um eine interne Quelle handelt hängt somit von dem Ursprung der Information ab.

Die zweite Definition unterscheidet zwischen Informationen, die öffentlich erkennbar sind und Informationen, die nicht öffentlich sind, sondern „verdeckt“ im Geiste ablaufen. Bei der Unterscheidung, ob es sich um eine externe oder um eine interne Quelle handelt, spielt somit der Zustand einer Information eine zentrale Rolle. Nach dieser Definition wäre beispielsweise die Unterscheidung zwischen dem, was eine Person in ihrer Vorstellung getan hat (nicht öffentlich, internal) und dem, was sie tatsächlich getan hat (öffentlich, external) dem Konzept der Realitätskontrolle zuzuordnen.

Auftretende Merkmale von Erinnerungen an reale Geschehnisse enthalten nach JOHNSON et al. (1993) tendenziell mehr perzeptuelle Einzelheiten (Farbe, Geräusche, ...), mehr kontextuelle Informationen (zeitlich, räumlich) sowie affektive und semantische Information und weisen auch eher auf unterstützende Gedächtnisinhalte hin, die ebenfalls perzeptuelle und kontextuelle Details beinhalten. Demgegenüber enthalten Erinnerungen an vorgestellte Ereignisse eher weniger der genannten Einzelheiten, jedoch dafür mehr Informationen über

kognitive Operationen, durch welche die Erinnerung entstanden ist, wie etwa Organisation, Behalten und Identifikation. Diese Unterschiede zwischen Gedächtnisinhalten auf der Basis von Vorstellungen bzw. der Realität sind ein direktes Spiegelbild bzw. stehen in direktem Zusammenhang mit den unterschiedlichen Bedingungen, die die Erinnerungen erzeugt haben (vgl. JOHNSON et al., 1993).

Neben der Forschung zum 'reality monitoring' von in der Regel alltäglichen Erinnerungsprozessen bzw. im experimentellen Labor-Setting erzeugten Gedächtnisinhalten sei an dieser Stelle zunächst einmal nur auf eine Studie von WARD & CARROLL (1997) hingewiesen, die von Versuchspersonen verwendete Realitätsüberwachungsstrategien bei sexuellem Missbrauch untersuchte und diese mit Strategien bei anderen Formen von Ereignissen verglich. Nach Befunden der Traumaforschung bzw. der Erinnerung an Traumata gibt es bedeutsame Unterschiede hinsichtlich der Art und Weise, wie Menschen traumatische und nicht-traumatische autobiographische Ereignisse erleben (Herman, 1992; Van der Kolk & Fisler, 1995; zit. nach WARD & CARROLL, 1997). In der Untersuchung von WARD & CARROLL (1997) beschrieben Versuchspersonen (im Alter zwischen 23 und 59 Jahren) inwiefern bzw. auf welche Weise sie wussten, dass vier verschiedene autobiographische Ereignisse (sexueller Missbrauch, ein andere Traumaerfahrung ohne Bezug, eine soziale Situation und ein vorgestelltes Ereignis) stattgefunden haben oder eben nicht. Im Ergebnis konnte festgestellt werden, dass Erfahrungen des sexuellen Missbrauchs ein besonderes, einzigartiges Antwortprofil hinsichtlich der Realitätsüberwachung aufzeigen, nach dem entsprechende Erinnerungen lebhafter sind als andere. Zudem finden sich hier weniger kontextuell-unterstützende Gedächtnisinhalte als bei anderen tatsächlichen Erinnerungen und mehr psychologisch argumentierende Antworten. Der Befund, dass emotionales Behalten ein Verblässen perzeptueller Einzelheiten verursacht, konnte für autobiographische Langzeiterinnerungen in dieser Studie nicht aufrechterhalten werden (vgl. WARD & CARROLL, 1997). Die genaueren Hintergründe und Implikationen werden im Diskussionsteil genauer beleuchtet.

1.1.6 Das Konzept der Quellendifferenzierung oder Quellenkontrolle (source monitoring)

Nach JOHNSON et al. (1993) lässt sich der allgemeinere Begriff der Quellenkontrolle auf die Unterscheidung zwischen mehreren externen und internen Erinnerungen subsumieren. Quellenkontrolle bezieht sich dabei auf diejenigen Prozesse, die für korrekte bzw. fehlerhafte Quellenattribution verantwortlich sind. Die Zuordnung von Erinnerungen zu verschiedenen Quellen wird als Funktion der zu beurteilenden Entscheidungsprozesse und Gedächtnisspuren verstanden, die aus der Gesamtheit erinnerter Attribute erschlossen werden müssen.

Das in späteren Revisionen des oben dargestellten Modells der Realitätsüberwachung eingeführte *Konstrukt der Quellenüberwachung (source monitoring)* stelle sich als übergeordnete kognitive Kompetenz dar. Neben der im Konzept der Realitätsüberwachung relevanten Diskrimination intern generierter gegenüber extern erfahrener Erinnerungen, lassen sich zusätzlich zwei Arten von Quellenkontrolle unterscheiden:

- (a) die Unterscheidung zwischen extern generierten Quellen (*'external source monitoring'*), also z.B. Erinnerungen an Aussagen, die entweder von Person A oder von Person B stammen
- (b) die Unterscheidung zwischen intern generierten Quellen (*'internal source monitoring'*), also z.B. zwischen Erinnerungen an Gedanken oder Ideen gegenüber ausgesprochenen Vorstellungen

Die beiden Arten der Quellenkontrolle spielen nach GREUEL (2001) im aussagepsychologischen Kontext allerdings nur eine marginale Rolle.

Hier geht es um die Unterscheidung zwischen verschiedenen Erinnerungsquellen, wobei im vorliegenden Zusammenhang insbesondere die Fähigkeit zur Unterscheidung zwischen verschiedenen externalen Quellen relevant ist (Dimension external-external). Dabei ist es von Bedeutung, dass sich Ereignisse, die ausschließlich aus mehreren externen oder nur internen Quellen stammen, in den für die Unterscheidung relevanten Attributen stärker überlappen als Ereignisse, die jeweils auf eine externe und auf eine interne Quelle zurückgehen. Demnach gilt es, Erinnerungen, die aus zwei oder mehr externen Quellen stammen, vor allem auch inhaltsspezifisch zu analysieren und nicht nur auf die Quantität bzw. das Ausmaß der Informationen bei der Quellenidentifizierung reduziert werden (vgl. JOHNSON, HASHTROUDI & LINDSAY, 1993).

Entsprechend der gegenwärtigen theoretischen Überlegungen und Annahmen basieren alle Arten von Quellenkontrolle auf einer Kombination von individuellen, bestimmten Erinnerungsmerkmalen und Entscheidungsprozessen, die im nachfolgenden Kapitel 'Einflussfaktoren und Fehler bei Quellenidentifikation und Quellenzuordnung' noch detailliert dargestellt werden.

In Bezug auf das Konzept des 'source monitoring' bei Kindern sind daneben vor allem auch zielorientierte Prozesse von Bedeutung aufgrund von Aktivitäten, nach denen Quellenunterscheidungen getroffen werden. Folgt man einer personen-zentrierten Perspektive, sind für das Verständnis von Genauigkeit und Fehlern im Rahmen der Quellenidentifikation auch der Bezug zu individuellen und allgemeinen Interessen (Zielen) von Bedeutung. Somit kann die Art und Weise internaler, auf das eigene Selbst bezogener Prozesse ebenfalls Hinweise auf die kognitive Leistungsfähigkeit in anderen Bereichen liefern (vgl. RATNER et al., 2000).

1.2 Psychologie der Zeugenaussage

1.2.1 Fehler und Unzuverlässigkeiten bei Zeugenaussagen – Modelle und Ursachen

Neben den Wahrnehmungs-, Erinnerungs- und (Re-)Konstruktionskompetenzen des Zeugen resultiert die Qualität einer Aussage immer auch aus dem kommunikativen Verhalten des Befragenden, stellt also somit ein unmittelbares, jedoch auch spezifisches Ergebnis eines *interdependenten Kommunikations- und Interaktionsprozesses* dar. Dabei zeichnen sich forensische Befragungen zwar durch ihren *institutionalisierten Charakter* aus, in dem Inhalte, Abläufe und die Struktur des Gesprächs häufig durch formal-strukturelle Rahmenbedingungen und Rollenverteilungen vorgegeben sind. Es finden sich sowohl auf einer sozialen als auch auf einer inhaltlichen Ebene in Form von mehrheitlich asymmetrisch geführter Kommunikation erhebliche Risiken sowohl bezüglich der Zuverlässigkeit als auch der Qualität der Aussage. Trotz der gegebenen normativen Restriktionen sollte ein Wechsel von symmetrischen zu asymmetrischen Kommunikationsebenen gewährleistet sein (vgl. GREUEL, 2001). Die Angaben eines Zeugen über ein beobachtetes Ereignis können aus unterschiedlichen Gründen ungenau oder falsch sein bzw. von verschiedenen Faktoren beeinflusst werden. Vergessensprozesse spielen hier oft eine Rolle, jedoch kann die Genauigkeit einer Aussage auch durch nachträgliche Informationen, die zum Zeitpunkt der Beobachtung eines Ereignisses gar nicht wahrgenommen werden konnten, beeinflusst werden. So bekommen die Hälfte der Versuchspersonen im klassischen Falschinformationsparadigma von Loftus verbal falsche, in Fragen eingebettete Informationen suggeriert, die sie zum Zeitpunkt der visuellen Beobachtung eines Ereignisses gar nicht wahrgenommen haben können; Versuchspersonen erzählen dann in ihrem späteren Bericht über das ursprüngliche Ereignis aufgrund dieser nachträglich irreführenden Informationen von Dingen, die sie erst zu einem späteren Zeitpunkt erfahren haben (GREUEL et al. 1998; MILNE & BULL, 2003).

Innerhalb mehrerer Theorien wurde versucht, den Falschinformationseffekt zu erklären: Loftus selbst (LOFTUS, 1979) geht in ihrer *Verschmelzungshypothese* davon aus, dass die ursprüngliche Erinnerung durch die „neue“ Information (Falschinformation) unwiderrufbar verändert oder überschrieben wird. Es könnte sich jedoch auch um Abruffehler handeln. Daher geht die *Koexistenz-Hypothese* (BEKERIAN & BOWERS, 1993) davon aus, dass die ursprüngliche Erinnerung unter günstigen Abrufbedingungen durchaus noch zugänglich sein kann und diese somit mit der verfälschten Version im Gedächtnis koexistiert. Schließlich könnte es auch sein, dass gar keine Gedächtnisveränderungen stattgefunden haben. Nach McCLOSKEY & ZARAGOZA (1985) handelt es sich hierbei lediglich um *Antwortverzerrungen* in den experimentellen Untersuchungen, die durch die verwendeten Versuchsanordnungen induziert wurden.

Grundsätzlich gilt bei der Befragung eines Zeugen, dass dieser nur das berichten soll, was er tatsächlich selbst gesehen oder gehört hat, unabhängig davon, was er selbst dabei gedacht hat oder was er aus anderen Quellen oder Gesprächen mit Dritten über das Ereignis erfahren hat (MITCHELL et al., 2003). Die Zuordnung eines Vorgangs zu seinem jeweiligen Ursprung ist für die Aussage eines Zeugen deswegen von zentraler Bedeutung, weil Zeugen häufig sowohl zwischen verschiedenen externalen, als auch zwischen externalen und internalen Quellen diskriminieren müssen. LINDSAY & JOHNSON (1989) sehen in fehlerhaften Quellenzuordnungsprozessen eine weitere mögliche Erklärung für die Suggestibilität von Zeugen durch nachträglich irreführende Informationen. Gemäß der *Quellenidentifikationshypothese* treten dann Fehler auf, wenn die irregeführten Versuchspersonen fälschlicherweise die Erinnerung an die Falschinformation als Erinnerung an das beobachtete Ereignis identifizieren.

REYNA et al. (2002) geben einen recht interessanten, umfassenden und mehrdimensionalen Überblick über die Entwicklung von für die forensische Praxis relevanten *'false memories'* bei Kindern, auf den an dieser Stelle nur kurz zusammenfassend hingewiesen werden kann und soll. Dabei führen die Autoren sechs zugrunde liegende Dimensionen an: Operationalisierung bzw. die Art der Messung, Entwicklungsverlauf, soziale Faktoren, individuelle Unterschiede, Vielzahl und Arten von Erinnerungen und Gedächtnis sowie deren Einschätzung sowie die Vielzahl der Vorgänge die *'false memories'* induzieren. Auf dem Hintergrund der Kritik einer mangelnden ökologischen Validität der bisherigen Studien sowie eines unzureichenden Gebrauchs insbesondere von bekannten Interviewtechniken, die eine bloße Zustimmung von tatsächlichen Gedächtnisveränderungen differenzieren, betonen Reyna et al. im Rahmen von Befragungen die Interaktion der beiden Aspekte im Sinne eines dynamischen Prozesses des Konstruierens von Fragetechniken und Erinnerungen. Demnach sollten Interviewer es vermeiden, Kinder dazu zu ermutigen, sich alternative Versionen von Ereignissen zu überlegen, vorzustellen oder zu interpretieren, insbesondere nicht durch wiederholte spezifische Fragen. Das Ziel einer Befragung sollte demnach neben der reinen Trennung zwischen wahren und falschen Inhalten die Unterscheidung von „zerfaserter“ Wahrheit, die Auffassung von Fragen sowie wesentlichen Erinnerungsmerkmalen und der wörtlichen „just-the-facts“-Wahrheit sein, welche für die institutionelle Verwaltung von Gerechtigkeit erforderlich ist (vgl. REYNA et al, 2002).

Trotz oder auch wegen der genannten Faktoren finden sich in Zeugenaussagen viele Arten, Formen und Aspekte von Fehlern, die im Folgenden beleuchtet werden sollen.

1.2.1.1 Einflussfaktoren und Fehler bei Quellenidentifikation und Quellenzuordnung

Die Unterscheidung zwischen internal und external generierten Erinnerungen erfolgt auf der Grundlage zweier basaler Strategien (JOHNSON & RAYE, 1981; JOHNSON et al., 1993):

- *Heuristische Strategien* auf der Basis von qualitativen Kriterien, die der Erinnerung selbst immanent sind (sog. Realitätsüberwachungskriterien)
- *Systematische Strategien* der Wirklichkeitskontrolle auf der Basis von Plausibilitätskriterien und subjektiven 'belief qualities'

Typischerweise werden Quellenurteile aufgrund heuristischer Überlegungen gefällt, da systematische Prozesse gewöhnlich weniger oft initiiert werden, mehr Zeit benötigen und anfälliger für Störungen sind. Umgekehrt können heuristische Prozesse eine Revision bei der Einordnung von Erlebnissen bewirken, die wegen unzureichender sensorischer Information akzeptiert und in ein allgemeines Schema eingeordnet werden (vgl. JOHNSON, 1988).

Den heuristischen Ausgangspunkt für das subjektive Urteil über den Wirklichkeitsstatus von Gedächtnisrepräsentationen bilden die qualitativen Merkmale der Erinnerung selbst. Demnach erhöht sich die Wahrscheinlichkeit, dass eine Erinnerung als external generiert eingestuft wird, wenn sie eine (relativ) hohe Anzahl von folgenden Aspekten, den sog. Realitätsüberwachungskriterien, enthält (Destun & Kuiper, 1999; Hashtroudi, Johnson & Chrosniak, 1990, Hashtroudi & Lindsay, 1993; JOHNSON & RAYE, 1981; zit. nach GREUEL, 2001):

- Kontextuelle Informationen (z.B. raum-zeitliche Angaben)
- Sensorische Informationen (z.B. Form, Farbe, Geruch, ...)
- Affektive Informationen (z.B. emotionale Reaktionen zum Zeitpunkt des Ereignisses)
- Semantische Informationen (z.B. bedeutungshaltige Details)

Die Attribution auf internal generierte Erinnerungen wird demgegenüber nicht nur in Ermangelung an Realitätsüberwachungskriterien wahrscheinlicher, sondern vor allem dadurch, dass Erinnerungen an internal Generiertes zusätzlich mehr Hinweise auf kognitive Operationen enthalten, die während der Elaboration weiterverarbeitet wurden, wie etwa Reflexionen oder Assoziationen (vgl. GREUEL, 2001).

Auf der anderen Seite ist es bei manchen Quellenentscheidungen auch nötig, sie langsamer und überlegter zu treffen und gegebenenfalls andere Erinnerungen in die Entscheidung miteinzubeziehen oder plausible Erklärungen heranzuziehen. Somit basieren systematische Strategien der Wirklichkeitskontrolle demgegenüber in erster Linie auf individuellen bzw. sozial verankerten Plausibilitätskriterien und Annahmen über Gedächtnisprozesse im Sinne des Metagedächtnisses, demzufolge die Wahrscheinlichkeit einer Attribution auf eine externe Erinnerungsbasis bei Vorliegen eines oder mehrerer der folgenden Kriterien steigen sollte (vgl. GREUEL, 2001):

- Kohärenz und Plausibilität (im Einklang mit natürlichen und logischen Gesetzmäßigkeiten bzw. Integration in domänenspezifisches Wissen)
- Rekonstruierbarkeit des Kontextes und weiterer Kontextinformationen
- Konfidenz/Sicherheit (Grundlage ist die Annahme, dass wirkliche Erinnerungen von subjektiv höherem Überzeugungsgrad sind. Im Falle von Unsicherheiten wird demzufolge eher auf internale oder aber außerhalb des eigenen Selbst liegende externale Quellen attribuiert („it-had-to-be-you-effect“; JOHNSON & RAYE, 1981). Dabei ist allerdings ein erhöhtes Risiko für Zuschreibungsfehler angelegt.

Systematisch getroffene Entscheidungen kommen jedoch seltener vor als heuristische, laufen langsamer ab und sind zudem anfälliger für Störungen (vgl. JOHNSON et al., 1993). Somit wird letztendlich ein gewichtetes und integratives Urteil über den Wirklichkeitsstatus einer Erinnerung gefällt. Die wenigen Studien zu den der Quellendiskrimination zugrunde liegenden Entscheidungsprozessen weisen darauf hin, dass die Realitätsüberwachung bzw. Quellendiskrimination ebenso von der Gedächtnisstärke und insbesondere auch von den spezifischen Abrufbedingungen mit beeinflusst wird (MARSH & HICKS, 1998). Die Genauigkeit von Quellenzuordnungen bzw. die Quellenidentifikation selbst im experimentellen Kontext ist dabei von mehreren Faktoren abhängig, wie z.B. Charakteristika der Frageformulierung oder Interviewtechnik, verschiedene Grade der verfügbaren und vorhandenen Information der aktivierten Erinnerungspfade, der Einschätzung der Menge an perzeptuellen Details, der Vertrautheit und Spezifität (vgl. JOHNSON & RAYE, 2000) und generell den jeweiligen Anforderungen an die entsprechende Situation. Daher basiert die Qualität der Quellenidentifikation fundamental auf den Informationen, die anfangs über das Ereignis aufgezeichnet wurden (JOHNSON & MULTHAUP, 1992). Somit ist ganz besonders die Qualität der Enkodierung eines Ereignisses oder einer Beobachtung von entscheidender Bedeutung. Alle Vorgänge - die das Ergebnis von spezifischen visuellen und kognitiven Prozessen sind, die während einer Wahrnehmung wirksam sind – und die die Person davon abhalten, die gesamten Kontextinformationen aufzunehmen, reduzieren somit die Enkodierung von potentiell relevanten Quelleninformationen. Bekannte Faktoren wie das Erregungsniveau, selektive Aufmerksamkeitsfokussierung oder bestimmte soziale Erwartungen beeinflussen ebenso wie die angesprochenen Quellen(-kombinationen) als Distraktoren unmittelbar die Entscheidungsprozesse (vgl. GREUEL, 2001). Allgemein lässt sich festhalten, dass geteilte oder selektive Aufmerksamkeit in bestimmten Situationen dazu führt, dass sich Personen nicht mit zusätzlichen Prozessen beschäftigen können, die nötig wären, um eine Information fest zu etablieren und für die Identifizierung der korrekten Quelle heranzuziehen. Die Zuordnung von Erinnerungen zu ihren Quellen kann unter gewissen Umständen anhand von recht undifferenzierten Informationen erfolgen und somit gegebenenfalls auch falsch sein. Es werden

jedoch spezifischere Informationen für die Quellenidentifikation herangezogen werden, wenn die richtige Zuordnung der Erinnerungen von größerer Bedeutung ist. Mann wird solange systematische Entscheidungsprozesse heranziehen, bis man genügend passende perzeptuelle, kontextbezogene, semantische und affektive Details erinnert hat, um eine richtige Quellenentscheidung treffen zu können.

Falls Zeugen oder Personen im Allgemeinen die Quelle einer Information nicht mehr erinnern können, greifen sie oftmals auf ihr früheres *schema-basiertes Wissen* zurück. So untersuchten SPANIOL & BAYEN (2002) spezifische Bedingungen, unter denen schema-basiertes Wissen im Rahmen von source-monitoring-Aufgaben angewendet wird und fanden, dass dies nur im Falle einer geringen asymptotischen 'alt-neu-Erkennung' von Items zutrifft (vgl. SPANIOL & BAYEN, 2002).

Wie bereits erwähnt beinhalten Erinnerungen an Ereignisse, die tatsächlich erlebt wurden, mehr perzeptuelle Details sowie mehr Kontextinformationen als Erinnerungen an Vorstellungen von Ereignissen. Die unterschiedliche Anzahl von solchen Erinnerungsmerkmalen gilt bei Entscheidungen der Realitätskontrolle als Basis dafür, ob es sich bei der Erinnerung um ein wirklich wahrgenommenes (externale Quelle) oder um ein vorgestelltes (internale Quelle) Ereignis handelt (vgl. HASHTROUDI, JOHNSON & CHROSNIAK, 1990; SUENGASS & JOHNSON, 1988; MUSSWEILER & NEUMANN, 2000). Dies konnte in einer Studie von JOHNSON et al. (1988) nachgewiesen werden: Die Versuchspersonen dieser Untersuchung sollten sich an vergangene aktuellere tatsächlich erlebte Ereignisse und an vergangene Träume oder Phantasien erinnern und diese jeweils anhand von einer Reihe von Gedächtnischarakteristika einschätzen. Sie beschrieben Ereignisse, die tatsächlich stattfanden, als reicher an Kontextinformationen und an sensorischen und perzeptuellen Einzelheiten. In einer separaten Studie sollten andere Versuchspersonen angeben, woran sie ein tatsächlich erlebtes Ereignis von einem vorgestellten Ereignis unterscheiden können. Auch für diese Probanden waren Erinnerungsmerkmale wie Kontextinformation und perzeptuelle Details ein Hinweis dafür, dass Ereignisse, die derartig genau beschrieben werden können, tatsächlich erlebt wurden.

In Realitätsüberwachungsaufgaben haben Abfragebedingungen und insbesondere die Kombination der zu diskriminierenden Quellen einen entscheidenden Einfluss auf die Performanz. So wird zum einen die perzeptuelle Ähnlichkeit zwischen den Erinnerungsquellen als einer der Hauptfaktoren für Quellenzuschreibungsfehler diskutiert (LINDSAY, 1990; LINDSAY et al., 1991; LINDSAY et al., 2004), zum andern kann durch systematische Manipulation der spezifischen Frageformulierung der Zuordnungsprozess gestört werden. Somit können nach GREUEL (2001) die der Quellenüberwachung zugrunde liegenden Urteils- und Entscheidungsprozesse nicht unabhängig von übergeordneten Kontextbedingungen verstanden und abgebildet werden.

HICKS et al. (2002) betonen, dass eine korrekte Quellenidentifikation und die dafür notwendigen Prozesse nicht unbedingt das Erinnern konkreter gespeicherter Details erfordert, sondern dass diese auch auf partielle, nur vage erinnerte Informationen zurückgreifen und qualitative Merkmale im Sinne eines subjektiven Gefühls des Erinnerns effektiv nutzen (vgl. HICKS et al., 2002).

In einer Untersuchung von MATHER et al. (2000) begingen Versuchspersonen mehr Quellenverwechslungen von der Art, die ihre Entscheidung unterstützten. Tendenziell wiesen sie - korrekter- wie inkorrektweise – von ihnen ausgewählten Möglichkeiten mehr positive Eigenschaften (Merkmale) zu als anderen und umgekehrt (*selektives Gedächtnis bzw. selektive Erinnerung von Informationen*). Einerseits dient dies der internalen Bestätigung bzw. dazu, sich in seiner Entscheidung gut zu fühlen und ein eventuelles Bedauern zu reduzieren, andere Optionen nicht gewählt zu haben oder falsche Entscheidungen oder Zuschreibungen getroffen zu haben. Andererseits ergeben sich daraus Probleme bezüglich der Richtigkeit und Genauigkeit sowie Zugänglichkeit von Gedächtnisinhalten und Erinnerungen wie auch hinsichtlich des Lernens aus vergangener Erfahrung.

Des weiteren zeigen einige Studien, dass die Genauigkeit von Quellenentscheidungen in speziellen Tests dadurch beeinflusst werden kann, dass die qualitativen Merkmale von Erinnerungen (perzeptuelle Details, semantische und affektive Informationen, Informationen über den räumlichen und zeitlichen Kontext usw.) verändert werden (LINDSAY et al., 1991; MARKHAM & HYNES, 1993). Somit führte in Untersuchungen von JOHNSON et al. (1981) die semantische Ähnlichkeit zwischen einer externen und einer internen bzw. zwischen zwei externen Quellen (LINDSAY, 1990; LINDSAY et al., 1991) zu größeren Schwierigkeiten und begünstigte Verwechslungen. Die Wahrscheinlichkeit für Fehler bei der Realitätskontrolle, also bei der Unterscheidung zwischen internalen und externalen Quellen, hängt somit von dem Ausmaß der Ähnlichkeit ab, das zwischen den Erinnerungen an das vorgestellte und an das erlebte Ereignis besteht (JOHNSON, 1988; JOHNSON & RAYE, 1981). Ebenso hängt auch die Wahrscheinlichkeit für die Verwechslung von zwei externalen Quellen (externe Quellenidentifikation) von der Ähnlichkeit ab, die zwischen den potentiellen Quellen existiert (HEKKANEN & McEVOY, 2002). So veränderten beispielsweise FERGUSON et al. (1992) die Ähnlichkeit von zwei gesprochenen Mitteilungen; es zeigte sich, dass sich die Genauigkeit der Quellenidentifikation verbesserte, als die Ähnlichkeit abnahm. Je größer die Übereinstimmung zwischen den Erinnerungen ist, desto eher kommt es zu einer Quellenverwechslung (JOHNSON et al., 1993).

Wie wichtig Erinnerungsmerkmale für die Quellenidentifikation sind, zeigt sich auch in der ungleichmäßigen Verteilung der Fehler, die bei Quellenentscheidungen auftreten. So ist beispielsweise die Wahrscheinlichkeit dafür, dass eine Versuchsperson fälschlicherweise angibt, ein Ereignis, über das sie nur gelesen hat, tatsächlich beobachtet zu haben, höher

als die Wahrscheinlichkeit dafür, dass die Versuchsperson angibt, sie hätte über ein Ereignis gelesen, obwohl sie es in Wirklichkeit tatsächlich beobachtet hat (BELLI et al., 1992). SMITH et al. (2001) verweisen im Rahmen der Forschung zu 'false memories' auf die Bedeutung von 'episodic confusion errors' sowie 'semantic confusion errors' (SMITH et al., 2001). Somit entwickelt man vermutlich bestimmte Vorstellungen über ein Ereignis, wenn man darüber liest. Beobachtet man dagegen tatsächlich ein Ereignis, ist es weniger wahrscheinlich, dass man sich vorstellt, man hätte etwas darüber gelesen (JOHNSON et al., 1993).

SUSSMAN (2001) untersuchte in ihrer Studie Altersunterschiede bei der Realitätsüberwachung interaktiver bzw. vorgestellter Ereignisse bei 4-, 8- und 12-Jährigen sowie Erwachsenen. Gleich im Anschluss sowie nach einer Woche erfolgten Gedächtnistests. 4-Jährige hatten größere Schwierigkeiten, vorgestellte Handlungen zu unterscheiden als 12-Jährige und Erwachsene. Bei den 4- und 8-Jährigen wurden vorgestellte Handlungen öfter mit tatsächlich ausgeführten verwechselt als umgekehrt. Aktivitäten, in denen der Teilnehmer selbst Ausführender der Handlung war, wurden besser unterschieden als solche, in denen dies jemand anderes war. Die Realitätsüberwachung nahm über die Zeit hinweg ab, speziell bei vorgestellten Dingen. Dies deckt sich mit verschiedenen anderen Befunden, als Kinder sofort bzw. dann wieder nach 2, 6 oder 8 Wochen zu Handlungen befragt wurden; so gaben diese während des zweiten Interviews weniger korrekte Antworten (Gordon et al., 1991, 1994; Jens et al., 1990; Parker, 1995; zit. nach SUSSMAN, 2001). Allerdings schnitten Kinder die ein zweites Mal befragt wurden besser ab als diejenigen, die nur einmal nach dem Delay befragt wurden (Parker, 1995). Insgesamt erweist sich eine unmittelbare Befragung als vorteilhaft, wenn das Interview zu einem späteren Zeitpunkt wiederholt wird. Daneben ist offensichtlich ein gradueller Entwicklungsanstieg hinsichtlich der Realitätsüberwachung von vorgestellten Ereignissen vom Vorschulalter bis in die Jugendzeit zu verzeichnen. Demnach scheint die Fähigkeit sich mit Realitätsüberwachung auseinanderzusetzen mit dem Alter von 11 oder 12 Jahren vollständig entwickelt (vgl. SUSSMAN, 2001). Inwieweit diese Befunde für lernbehinderte Jugendliche oder generell Menschen mit Intelligenzminderung gelten oder übertragbar sind, bleibt auch aufgrund der Individualität und Komplexität der Störung zumindest kritisch nicht gänzlich klar zu beantworten.

Hinsichtlich entwicklungsbedingter Faktoren stellten z.B. MATHER et al. (1999) fest, dass ältere Menschen Schwierigkeiten beim Erinnern von Kontextinformationen haben und weniger dazu fähig sind, kontextuelle Eingaben adäquat zu nutzen. Zudem haben ältere Personen Defizite bei der Erinnerung bezüglich der Modalität der Präsentation (verbal durch den Versuchsleiter oder visuell via Dia- oder Filmprojektoren). Ebenso hängen die Forschungsergebnisse von der Art der Aufgabe ab. So stellen für Kinder (aber auch Erwachsene) einige Quellenkontrollsituationen keinerlei Schwierigkeiten dar, andere wiederum jedoch durchaus. So neigen nach FOLEY & JOHNSON (1985) Kinder viel eher dazu, die Erinnerungen für

Vorstellungen mit Erinnerungen für wirkliche Tätigkeiten zu verwechseln (internale Quellenkontrolle), aber es zeigten sich weder Unterschiede bei der Realitätskontrolle noch bezüglich externaler Quellenkontrolle. Auch LINDSAY et al. (1991) fanden bei allen Arten von Quellenzuordnungen, dass Altersunterschiede umso eher deutlich werden, je ähnlicher sich die zu unterscheidenden Quellen sind. Somit werden andere mit der Herkunft verbundene Quellen umso wichtiger, je ähnlicher verschiedene Quellen auf einer Dimension sind. Weiterhin fanden einige Studien Unterschiede bei Personen mit hoher bzw. niedriger Vorstellungskraft. So folgerten MARKHAM & DOBSON (1993), dass Menschen mit hohen visuellen Fähigkeiten eher dazu neigten, ihre intern generierten Bilder auf präsentierte Filminformationen zu beziehen, weil deren vorhandene Vorstellungsbilder viele visuelle Details beinhalteten. Umgekehrt generierten diejenigen Versuchspersonen mit niedriger Vorstellungskraft interne Bilder, die weniger perzeptuelle Qualitäten enthielten und somit weniger dazu neigten, ihre Gedächtnisbilder auf die visuelle Quelle zu attribuieren. Insgesamt lassen die Ergebnisse vermuten, dass eine klarere mentale Repräsentation zu Entscheidungen bei Quellenzuordnungen beiträgt, indem sie eine Vorstellung eher als „Erinnerung“ etikettiert und weniger eine andere Quelle dafür verantwortlich macht. Die vorliegenden Befunde betonen dementsprechend, dass individuelle Faktoren hinsichtlich der mentalen Repräsentationsfähigkeit bei Studien zur Erinnerungsleistungen und der Unterscheidung von Informationen und deren Herkunft durchaus von Bedeutung sind und berücksichtigt werden müssen (WINOGRAD et al., 1998). Auch MITCHELL et al. (2003) betonen die entwicklungsbezogenen Unterschiede, die bei der Quellenidentifikation auftreten. So haben beispielsweise ältere Erwachsene größere Schwierigkeiten darin, zu entscheiden, welcher von zwei Sprechern eine Rede gehalten hat, in welcher Liste ein spezielles Item präsentiert wurde oder ob sie etwas gelesen oder selbst generiert haben (vgl. KAUSLER, 1994; SPENCER & RAZ, 1995). Die Ergebnisse von MITCHELL et al. (2003) deuten darauf hin, dass ältere Erwachsene schlechter oder gar nicht in der Lage sind, hilfreiche, diagnostische Quelleninformationen zu nutzen (z.B. perzeptuelle Details oder zeitliche Informationen) und dass sich deren Vertrauen in ihre falschen Erinnerungen in einem Übermaß an semantischer Information widerspiegelt bzw. sie sich in erster Linie darauf stützen (vgl. MITCHELL et al., 2003). Neuere Forschungen führen hauptsächlich zwei Gründe für die schlechtere Quellenidentifikationsleistung von älteren Erwachsenen an: Zum einen scheinen ältere Menschen manchmal Schwierigkeiten dabei zu haben, gleichzeitig verschiedene Hinweisreize, wie perzeptuelle, kontextbezogene, affektive oder semantische Details für eine Quelle heranzuziehen (BAYEN & MURNANE, 1996); zum anderen stellt die Zuordnung von Informationen zu ihren Quellen für ältere Erwachsenen vor allem dann ein Problem dar, wenn eine große Ähnlichkeit zwischen den beiden Quellen besteht (HENKEL et al., 1998). FERGUSON et al. (1992) konnten in einer Untersuchung beispielsweise nachweisen, dass es älteren Erwachsenen schwerer als

jüngeren Erwachsenen fällt, zu entscheiden, welcher von zwei weiblichen Sprechern eine Rede gehalten hat. Es zeigten sich zwischen den beiden Altersklassen jedoch keine Unterschiede in der Unterscheidung zwischen einem weiblichen und einem männlichen Sprecher als mögliche Quellen.

Johnson et al. (1979; zit. nach COHEN, 1996) untersuchten die Annahme, dass Menschen die mental besonders gut dazu in der Lage sind, lebendige und detaillierte visuelle Bilder zu konstruieren, weniger gut bei Einschätzung von Realitätsüberwachung sind, da die Lebendigkeit ihrer selbst generierten Bilder qualitativ external hergeleiteten Erinnerungen ähnlich ist und wird. Umgekehrt sollte Personen mit schlechter Vorstellungskraft ihre verschwommenen, unklaren Bilder und Vorstellungen besser von tatsächlichen Erinnerungen unterscheiden können. Die Experimente bestätigten diese Annahme mit der Schlussfolgerung, dass ein gutes Vorstellungsvermögen Schwierigkeiten bei der Realitätsüberwachung bereitet (vgl. COHEN, 1996). In diesem Zusammenhang (der Quellendifferenzierung in Abhängigkeit vom bildhaften Vorstellungsvermögen) ist ebenfalls die schon erwähnte Untersuchung von DOBSON & MARKHAM (1993) zu nennen. Dabei sahen Psychologiestudenten (im ersten Jahr) einen Film über ein Verbrechen und erhielten zudem eine schriftliche Beschreibung des Geschehens, die Informationen enthielt, die im Film nicht zu sehen waren. In einer Quellenidentifikationsaufgabe erwiesen sich Personen mit hoher und niedriger Vorstellungskraft als vergleichbar gut bei der Wiedererkennung zuvor präsentierter Items. Personen mit hohem Vorstellungsvermögen schnitten allerdings bei der Quellenunterscheidung von Items, die sich auf die schriftliche Beschreibung bezogen, schlechter ab. Diese Befunde weisen ebenfalls auf mögliche Enkodierungs- und Behaltensunterschiede aufgrund eines lebendigen und bildhaften Vorstellungsvermögens hin (vgl. DOBSON & MARKHAM, 1993). Inwieweit Lernbehinderte eher über reichhaltiges Vorstellungsvermögen verfügen oder nicht und sich dies auf die Qualität von Gedächtnisleistungen auswirkt, bleibt zu diskutieren.

Diese *Muster selektiver Defizite* sollten jedoch nicht als Beweis dafür angesehen werden, dass verschiedene Quellenzuordnungsaufgaben unterschiedliche Erinnerungs- oder Entscheidungsprozesse umfassen oder beanspruchen. Wenn allerdings zu unterschiedlichen Aufgaben bezüglich der Quellenidentifikation andere Erinnerungsmerkmale herangezogen werden (z.B. sind kognitive Vorgänge weitaus wichtiger, um interne von externen Quellen zu unterscheiden als bei der Diskrimination von zwei externen Ereignissen), können beobachtete Fehlleistungen bei defizitären Gruppen dazu beitragen, diejenigen Aspekte zu identifizieren, die teilweise nicht enkodiert, reaktiviert oder bei Entscheidungen herangezogen werden. Beide Arten von Entscheidungsprozessen, sowohl die heuristischen als auch die systematischen, können von bestimmten Vorurteilen, Vermutungen oder gegenwärtigen Zielen beeinflusst werden. So wird man beispielsweise eine nur vage vorhandene Erinnerung an ein Ereignis für eine Erinnerung an ein vorgestelltes Ereignis halten, weil man vermutet, dass

man sich besser an das Ereignis erinnern könnte, wenn man es tatsächlich erlebt hätte. Wie jedes zielgerichtete Vorgehen hängen auch die Quellenentscheidungen von motivationalen und sozialen Faktoren ab. Unter manchen Umständen wird die Entscheidung für eine bestimmte Quelle vorsichtiger gefällt und die Kriterien für ein Urteil strenger gesetzt als unter anderen Umständen. So wird man bei einer Aussage vor Gericht bei der Zuordnung von Informationen zu Quellen sorgfältiger vorgehen als bei der Erinnerung an bestimmte Ereignisse aus dem sozialen Umfeld. Mit dem Begriff „sorgfältig“ ist dabei gemeint, dass das gewählte Kriterium möglichst eng gewählt wird und zur Entscheidungsfindung sowohl heuristische als auch systematische Prozesse eingesetzt werden (vgl. JOHNSON et al., 1993).

Im aussagepsychologischen Kontext hat die Diskrimination zwischen verschiedenen externen Quellen nur nachrangig mit der Qualität einer Aussage zu tun, als vielmehr mit der Aussagezuverlässigkeit, wobei hier wiederum die Abschätzung möglicher Suggestionseffekte durch Dritte relevant sind. Die bis dato eher defizitäre Forschungslage zur Unterscheidung externer Quellen konzentrierten sich dabei vor allem auf visuelle und akustische Reize (MARSH & HICKS, 1998). Der auch diesbezüglich auftretende Effekt und Einfluss von Falschinformationen ist ein relativ häufig vorkommendes und resistentes Phänomen, dessen zugrunde liegende Mechanismen noch nicht klar bewiesen sind. Neben der mittlerweile nicht mehr aufrechtzuerhaltenden These von Loftus (1979), die von einer Veränderung der Gedächtnisspur über die Originalinformationen aufgrund von Löschung existierender und Ersetzung durch inkorrekte Iteminformation ausgeht, bestehen verschiedene alternative Interpretationen. Somit kann beispielsweise vermutet werden, dass die relevante Information nicht während des Originalereignisses enkodiert wurde, sondern der Zeuge die falsche Information nach der Wahrnehmung akzeptiert und bei einer Befragung eine entsprechende Aussage macht. Weiterhin wäre eine *Koexistenz der Original- und Falschinformation* denkbar, wobei die korrekten Details während der Abrufphase aufgrund der widerstreitenden fehlerhaften Aspekte nicht mehr zugänglich sind (vgl. auch HYMAN & PENTLAND, 1996). Diese Vermutungen werden durch Befunde gestützt, die besagen, dass durch die Wiederherstellung des ursprünglichen Wahrnehmungskontextes beim Abruf die Originalereignisse besser erinnert und Fehler reduziert werden (vgl. BEKERIAN & BOWERS, 1983). Es stellt sich weiterhin die Frage, ob und inwieweit irreführende und/oder falsche Bemerkungen/Informationen das Gedächtnis für visuelle Informationen beeinträchtigen. Gemäß der Quellenidentifikationshypothese treten Fehler in der Testphase dann auf, wenn die fehlgeleiteten Personen fälschlicherweise Erinnerungen an die Falschinformation als Erinnerung an das beobachtete Ereignis identifizieren. Die gefällten Urteile werden durch die subjektiven Entscheidungsprozesse und Kriterien beeinflusst. Diese Maßstäbe, die man zur Einordnung einer Erinnerung zu einem bestimmten Ursprung heranzieht, sind z.B. der Verwendungs-

zweck der Erinnerung, die individuell vorhandenen Einstellungen und Vorurteile oder die Schlüssigkeit der zu erinnernden Zusammenhänge und Tatsachen (vgl. JOHNSON, 1988).

Die Ergebnisse der Untersuchung von ZARAGOZA & LANE (1994) belegen, dass Versuchspersonen unter bestimmten Umständen zu der Überzeugung gelangen, sie hätten gewisse Dinge gesehen, die ihnen jedoch nur suggeriert wurden (*source misattribution effect*). Allerdings variiert die Größe des gefunden Effekts deutlich und es kann nicht daraus gefolgert werden, dass falsche Quellenzuschreibungen zwangsläufig die Konsequenz des Ausgesetztseins gegenüber Suggestionen darstellen (vgl. ZARAGOZA & LANE, 1994).

LINDSAY et al. (2004) untersuchten in ihrer Studie den Effekt des Ausmaßes der konzeptuellen Ähnlichkeit zwischen einem beobachteten Ereignis und einer zusätzlichen Erzählung auf die Suggestibilität von Augenzeugen. Es zeigte sich unter anderem, dass Versuchspersonen manchmal Details einer narrativen Beschreibung (Erzählung) eines Ereignisses in ihre Schilderungen eines anderen visuellen Geschehens einbauen bzw. einfließen lassen. Wenn die Erzählung das visuelle Geschehen selbst beschrieb, waren die Intrusionsraten sogar noch höher. Zwar sind nicht alle Zeugen Fehlinformationen im Zusammenhang mit Aussagen über beobachtete Ereignisse ausgesetzt, jedoch verfügen praktisch alle Zeugen über vergangene Erfahrungen, die Details beinhalten, die in irgendeiner Form auf im forensischen Setting eventuell gestellte Fragen passen. Zusammen mit den Befunden früherer Arbeiten folgern LINDSAY et al. (2004), dass es von mehreren Faktoren abhängt, ob die Veränderung der Ähnlichkeit von Quellen die Suggestionsanfälligkeit beeinflusst: erstens müssen die Abfrage-Instruktionen entsprechend gestaltet sein, zweitens muss die Abfrage es den Teilnehmern erlauben, nahe liegende Einzelheiten zu schildern und drittens muss die Veränderung der Quellenähnlichkeit im Kontext anderer Aspekte der Aufgabe ausreichend stark sein. Demgegenüber konnte jedoch bei Kindern auch festgestellt werden, dass bei Unsicherheit über den Ursprung einer Handlung besseres Lernen einer Aufgabe erreicht werden konnte (Ratner et al., 2000), dass Kinder mit einem Verständnis für Gedächtnis und guten Erinnerungen verschiedener Quellen Quellenverwechslungen begingen (Welch-Ross, 1999) und dass Kinder, die Erinnerungen über vielfältige Vorfälle verwechselten, kaum völlig falsche Informationen berichteten (Powell et al, 1999; zit. nach ROBERTS, K. P., 2002). Die Kompetenz von Kindern in anderen kognitiven Fähigkeiten geht also auf Kosten ihrer Quellenidentifikation, was aber nicht bedeutet, dass entsprechende Berichte ungenau oder falsch sein müssen (vgl. ROBERTS, K. P., 2002).

Insgesamt kann aufgrund der verschiedenen Untersuchungen und Befunde festgehalten werden, dass es wohl schwieriger ist Gedächtnisinhalte und Erinnerungen zu differenzieren, die beide entweder external oder internal generiert sind als eine Erinnerung, die sowohl eine externale als auch eine internale Quelle aufweisen (vgl. DOBSON & MARKHAM, 1993).

LINDSAY (1994) fasst eine Reihe von Faktoren zusammen, die die Art an der die Quelle einer Erinnerung identifiziert wird, beeinflussen:

- Die Menge und Struktur quellenrelevanter Information im aktivierten Erinnerungsverzeichnis
- Die Menge und Struktur des zugänglichen Wissens über die charakteristischen Eigenschaften der Quelle
- Das Ausmaß bis zu dem die Attribute des fraglichen Gedächtnisinhalts einmalig spezifisch für die Erinnerungen dieser Quelle sind
- Die Stringenz und Angemessenheit der Entscheidungsprozesse und verwendeten Kriterien während des Erinnerns

In Analogie zur Wahrnehmung können die genannten Faktoren weiter erläutert und verständlich gemacht werden: Wahrnehmungsfehler sind relativ wahrscheinlich, wenn die Stimuli vermindert oder unbekannt sind, wenn die Ziele den Distraktoren ähnlich sind und wenn die Entscheidungskriterien gering oder unpassend sind. Bleiben alle anderen Faktoren gleich, sind Quellenidentifikationsfehler auch dann wahrscheinlich, wenn Erinnerungen unklar oder unvollständig sind, wenn die Quelle unbekannt ist, wenn mehr als eine Quelle charakteristisch zu Erinnerungen führt, die dem fraglichen Gedächtnisinhalt ähnlichen sind oder wenn die Zuschreibung eher schnell und automatisch als aufgrund sorgfältiger Überlegung erfolgt (vgl. LINDSAY, 1994).

Somit kann davon ausgegangen werden, dass die subjektiven Beurteilungen von Erinnerungen je nach situativen und individuellen Kontextbedingungen variieren können (vgl. Detsun & Kuiper, 1999; zit. nach GREUEL, 2001).

1.2.1.2 Formen und Ursachen suggestiver und nachfolgender Beeinflussung

Für den Bereich der Psychologie der Aussage und der Glaubhaftigkeitsbegutachtung ist vor allem die *Suggestion durch Befragungseffekte (interrogative Suggestibilität)* bedeutsam, wobei sich grundsätzlich zwei Forschungsansätze unterscheiden lassen. Zum einen ein differentialpsychologischer Ansatz, dessen Focus auf der (suggestiv zu beeinflussenden) Person liegt im Sinne einer Identifizierung eines 'personalen traits' bzw. Korrelaten personaler Suggestibilität. Zum anderen ein allgemeinpsychologisch orientierter Ansatz, der sich unmittelbar mit vor und/oder in der Befragungssituation selbst angelegten Wirkfaktoren und der Variation von situativen Merkmalen des Befragungskontextes selbst befasst (vgl. GREUEL, 2001). BRUCK & MELNYK (2004) geben einen ausführlichen Überblick im Sinne einer Metaanalyse aus 69 Studien, die den Zusammenhang von demographischen, kognitiven sowie psycho-sozialen Faktoren und Suggestibilität bei Kindern zusammenfasst. Im

Ergebnis zeigten sich über die zahlreichen Untersuchungen hinweg vor allem bei kognitiven Faktoren, sprachlicher Ausdrucksfähigkeit sowie Kreativität ziemlich gut übereinstimmende Zusammenhänge mit Suggestibilität. Im Bereich der psycho-sozialen Faktoren wurde die größte Korrelation hinsichtlich der Maße von Selbst-Konzepten, mütterlicher Bindung und der Eltern-Kind-Beziehung allgemein gefunden (vgl. BRUCK & MELNYK, 2004). Dem an kindlicher Suggestibilität interessierten Leser sei diese Zusammenstellung und Überblick über die Suggestionforschung in jedem Fall empfohlen.

Im Folgenden sollen verschiedene Formen suggestiver Verhaltensweisen jeweils kurz erläutert werden (vgl. KÖHNKEN, 1999B; VOLBERT, 1997; GREUEL, 2001):

Induzierung von Stereotypen (unspezifische Suggestion):

Um im Rahmen der Zeugenvernehmung die bei Kindern oder behinderten Menschen vermuteten Ängste oder Blockaden zu lösen, werden suggestive Prozesse oftmals durch direkte (in Form von negativen Äußerungen über bestimmte Personen) wie auch indirekte Aktivierung (bestimmte Assoziationen durch Anreicherung der Aussagen der Zeugen) von Personenstereotypen induziert. Diese stellen oft die Grundlage für andere Formen suggestiver Beeinflussung in Kombination mit diesen Stereotypisierungen.

SHERMAN & BESSENOFF (1999) untersuchten den Gebrauch stereotyper Erwartungen als Quellenhinweise für biographische Erinnerungen. Im Ergebnis wiesen die Versuchspersonen einer Zielperson eher stereotype Verhaltensweisen zu, wenn die Quelle des ursprünglichen Verhaltens schwer zugänglich war (Quellenverwechslung) und Erinnerungsprozesse gestört waren. Deshalb verlassen sich erinnernde Personen, wenn die Erinnerung episodischer Details unterbrochen ist, möglicherweise eher auf semantisches Wissen, um Gedächtnishalte zu interpretieren. Dabei erweist sich der Gebrauch von Stereotypen nicht nur während des Enkodierens sozialer Information als effizient, sondern auch während des Behaltensprozesses (vgl. SHERMAN & BESSENOFF, 1999).

Nachträgliche Informationen:

Häufig werden einem Kind oder Zeugen generell in Vernehmungen Detailinformationen vorgegeben, die bis zu diesem Zeitpunkt von dem Befragten selbst noch nicht erwähnt wurden, was exakt dem Paradigma des „Falschinformationseffekts“ entspricht, zu dem inzwischen eine Vielzahl von Studien vorliegt und dessen Konzept schwerpunktmäßig auf kognitiven Faktoren und gedächtnispsychologischen Prozessen basiert. Dennoch gehen diese Suggestionseffekte eng mit sozialer Einflussnahme einher. Da jedoch leider in den klassischen Standarduntersuchungen zum Falschinformationseffekt fast ausschließlich Wiedererkennungseleistungen von den Versuchspersonen gefordert werden, haben diese Studien nur eingeschränkten Erklärungswert für die angewandte Psychologie der Aussage,

da dort vielmehr freie Reproduktionen komplexer Erlebnisse von Bedeutung sind. Trotzdem der Effekt empirisch gut belegt ist, besteht dennoch Dissens bezüglich der zugrunde liegenden Mechanismen, wobei neben Quellenzuordnungsfehlern (JOHNSON et al., 1993) zunehmend sozialpsychologische Erklärungsansätze favorisiert werden (vgl. GREUEL, 2001).

Wiederholung von Fragen:

Verschiedene Untersuchungen (z.B. CASSEL, ROEBERS & BJORKLUND, 1996; POOLE & WHITE, 1991) haben gezeigt, dass gerade Kinder dazu neigen, ihre erste - möglicherweise richtige - Antwort zu ändern und eine andere Antwort zu geben, wenn ihnen in kurzem zeitlichem Abstand wiederholt dieselbe (geschlossene) Frage gestellt wird. Dieser Effekt tritt dagegen nicht oder in sehr viel geringerem Maß auf, wenn offene Fragen wiederholt gestellt werden. Dieses Phänomen wird nach Poole & White damit erklärt, dass Befragungen nicht einfach Formen von Wissens- oder Gedächtnistests sind – zumindest nicht explizit – sondern Konversation, die nach gelernten Konversationsregeln geführt wird, auch wenn dies in der forensischen Praxis, aber auch in der wissenschaftlichen Forschung oftmals nicht ganz so ablaufen kann oder soll. Da gerade Kinder – wie auch behinderte Menschen - auf offene Fragen relativ wenig Information produzieren, besteht hier schnell eine starke Tendenz, auf direkte geschlossene Fragen zurückzugreifen. Dabei impliziert das Wiederholen einer Frage oft, dass die erste Antwort unzulänglich oder falsch war, besonders wenn dies zudem mit einer negativen Rückmeldung verbunden ist. Dies entspricht einer typischen, nicht nur alltäglichen, sondern leider auch in der forensischen oder polizeilichen Vernehmungspraxis üblichen Kommunikationssituation zwischen Erwachsenen und Kindern, bei der Erwachsene häufig Fragen stellen, auf die sie selbst die richtige Antwort kennen oder diese zumindest vermuten und es nun Aufgabe des Kindes ist, diese „richtige“ Antwort zu finden. Eine mögliche Lösung, diese Signalwirkung zu reduzieren, besteht darin, eine Fragewiederholung damit zu begründen, dass der Interviewer etwas nicht richtig verstanden hat, um nicht den Eindruck nahe zulegen, das Kind habe einen Fehler gemacht (vgl. VOLBERT, 1997; KÖHNKEN, 1999B).

Wiederholte Befragungen (verschiedene Interviews zum selben Sachverhalt):

Für Erwachsene ist empirisch inzwischen gut belegt, dass sich Wiederholungen (rehearsals) des Erinnerungsmaterials positiv auf die Gedächtnisleistung auswirken (z.B. Cohen, 1996; zit. nach GREUEL, 2001). Für kindliche Zeugenaussagen gestaltet sich dies nicht so klar bzw. problematisch. Dabei ist jedoch zu beachten, dass die potentiellen Verfälschungseffekte auf kindliche Gedächtnisreproduktionen in erster Linie auf das *Moment* der unsachgemäßen, suggestiven Befragungstechnik und erst nachrangig auf den Umstand der Befragungswiederholung als solchen zurückgeführt werden kann. So belegen Studien sogar Gegenteil-

ges, dass wiederholte suggestionsfreie Befragungen durchaus die Erinnerungsleistung von Kindern begünstigen können (vgl. FIVUSH & SCHWARZMÜLLER, 1995; POOLE & WHITE, 1995).

Wenn im Wesentlichen offene Fragen gestellt werden, können wiederholte Befragungen unter Umständen dazu beitragen, Vergessensprozesse zu reduzieren (vgl. Poole & White, 1995; Warren & Lane, 1995; zit. nach VOLBERT, 1997). Werden wiederholte Befragungen dagegen in suggestiver Form geführt, kommen einige Untersuchungen (vor allem Ceci, Loftus et al., 1994; Bruck et al., 1995; zit. nach VOLBERT, 1997) zum dem Ergebnis, dass diese einen erheblich negativen Effekt auf die Aussage haben. In Untersuchungen von Warren & Lane (1995) und Volbert & Dippold (1998) wurde deutlich, dass neben der Wiederholung insbesondere dem Zeitpunkt der Präsentation der Falschinformation eine große Rolle zukommt. Demzufolge haben möglicherweise einzelne späte suggestive Befragungen stärkere negative Effekte als wiederholte frühe und wiederholte späte suggestive Befragungen haben besonders ausgeprägte Suggestionseffekte. Während man lange davon ausging, dass der spontane Bericht bzw. Antworten auf offene Fragen in hohem Maß richtig sind und von Suggestionen unbeeinflusst bleiben, zeigt sich zunehmend, dass suggerierte Informationen auch auf eine offene Erzählaufforderung oder auf offene Fragen produziert werden, wenn mit wiederholter suggestiver Beeinflussung operiert wird (vgl. Bruck et al., 1995; Ceci, Loftus et al., 1994; Poole & White, 1995; zit. nach VOLBERT, 1997). Darüber hinaus scheinen die Aussagen im Verlauf von wiederholten Befragungen zunehmend detaillierter zu werden.

Weiterhin zeigten Mitchell & Zaragoza (1996) und Zaragoza & Mitchell (1996), dass bei erwachsenen Versuchspersonen das Ausmaß falscher Quellenzuordnungen von Falschinformationen bei wiederholter Falschinformationspräsentation höher liegt als bei einmaliger (zit. nach VOLBERT, 1997). Bei Kindern (oder lernbehinderten Menschen) könnte dieser Effekt sogar noch größer sein, wobei auch der Zeitpunkt eine Rolle spielt. Parker (1995) konnte zeigen, dass die Zahl der Falschzuordnungen höher lag, wenn die Befragung nicht in unmittelbarem Anschluss an die Informationspräsentation erfolgte, sondern eine Woche später. Mitchell & Zaragoza (1996) weisen außerdem auf die Bedeutung der kontextuellen Variabilität hin, wonach Quellenzuordnungsfehler nach wiederholter Falschinformationspräsentation noch zunehmen, wenn die Suggestion in verschiedenen Modalitäten präsentiert werden (Video, gedruckt,...).

Konformitätsdruck:

Dabei werden Kinder wie Erwachsene damit konfrontiert, dass z.B. Angehörige der 'peer-group' oder andere nahe stehende Vertrauenspersonen die Veridikalität der inhaltlich falschen Suggestion bereits bestätigt haben. Daneben stehen bei der Technik des 'Coaching'

suggestionsbedingte Quellenverwechslungen bzw. Übertragungseffekte im Mittelpunkt des Forschungsinteresses.

Systematische Konditionierung:

KÖHNKEN (1999B) weist darauf hin, dass Konditionierungen in Form von positiver wie negativer Verstärkung erwarteter bzw. nicht erwarteter Antworten und Ignorieren fatale Auswirkungen auf die Aussage eines Zeugen haben können, da eine Befragung auch immer eine Lernsituation darstellt, in der die befragte Person aus den Reaktionen des Fragenden lernt, welche Reaktionen von ihr erwartet werden. Dies kann zu einer Verhaltensformung führen, die schließlich in einer vermehrten Produktion der verstärkten Inhalte resultiert. Der durch die ständig wiederholten Befragungen ausgeübte Druck stellt eine aversive Situation dar, die erst beendet ist oder gelindert wird, wenn die erwarteten Angaben gemacht werden.

Aufforderung zu Konfabulationen:

Dies stellt eine besonders folgenschwere Suggestionstechnik dar, ebenso wie das Deuten von Träumen oder Zeichnungen sowie die Imagination fiktiver Ereignisse (vgl. KÖHNKEN, 1999B). In Kombination mit anderen Suggestionstechniken können auf diese Weise Gedächtnisinhalte implantiert werden, die später nur noch schwer von realen Erinnerungen zu unterscheiden sind (Quellenzuschreibungsfehler; LINDSAY, 1994).

Im Rahmen von *Suggestionen bei neutralem Ausgangserlebnis* kann nach GREUEL (2001) nach Bezugnahme auf verschiedene Forschungsergebnisse zusammenfassend festgestellt werden:

- Angaben von Kindern im freien Bericht bzw. auf Anstoßfragen sind insgesamt sehr genau, wenn auch unvollständig. Durch Anstoßfragen wird die Vollständigkeit von (Kinder-) Aussagen generell erhöht, wobei die zusätzlich generierten Aussagen weitestgehend genau sind.
- Suggestive geschlossene Ja/Nein-Fragen sind für (Vorschul-)Kinder insofern problematisch, da hier eine Ja-Sage-Tendenz besteht. Unabhängig vom Alter treten bei Kindern Verfälschungseffekte auf der Basis multipler Suggestivfragen in erster Linie in Bezug auf periphere und deskriptive Details des persönlichen Interaktionspartners auf, hingegen nur selten in Bezug auf zentrale Handlungsaspekte.
- Die als 'Coaching' bezeichnete Suggestionstechnik sowie der entsprechende Effekt tritt auf der Aussageebene primär in Bezug auf schemakonsistente Details auf, kann jedoch deutlich reduziert werden, wenn Kinder explizit nach der Quelle bzw. Basis ihrer Antwort (selbst erlebt vs. gehört) befragt werden.

- Vorinformierte Befrager stellen zwar mehr hypothesenkonforme (geschlossene) Fragen als uniformierte Befrager (interviewer bias), allerdings geht eine einseitige Hypothesenbildung nicht unbedingt mit einem Anstieg irreführender Suggestivfragen einher. Jedoch steht insbesondere bei dem Kind unvertrauten Interviewern eine entsprechende Voreinstellung des Befragers mit einer Reduktion der Aussagegenauigkeit in Zusammenhang, primär in Form von mehr Auslassungsfehlern und Hinzufügen von (nicht suggestiv vorgegebenen) Details und Fehlern bezüglich der Reihenfolge von Ereignissequenzen.
- Mehrmalige Befragungen führen nicht per se zu einer Reduktion der Aussagezuverlässigkeit. In Folgebefragungen – und damit mit zunehmendem Erinnerungsintervall – ist zwar mit einer Abnahme des Aussageumfangs und der generellen Aussagegenauigkeit zu rechnen, die aber nicht zwangsläufig mit einer Erhöhung suggestionsbedingter Aussagefehler einhergeht. Durch Befragungswiederholungen kann es auch zu einer Abnahme der Aussagefehler oder sogar zu Reminiszenz und damit (zuverlässigen) Aussageerweiterungen bis hin zur Hypermnesie kommen. Selbst nach suggestiver Vorbefragung ist im freien Bericht des Folgeinterviews weitestgehend mit zuverlässigen Angaben zu rechnen.

Auf der Basis von Suggestivfragen konnten bedeutsame Suggestionseffekte nur in seltenen Ausnahmefällen nachgewiesen werden. Die so erhaltenen Angaben der Kinder erwiesen sich durchgängig als detailarm in Bezug auf den narrativen Handlungskern, wonach nach aufgrund der derzeitigen empirischen Befundlage eine sexuelle Umdeutung und Fehlrepräsentation originär neutraler Erlebnisinhalte allein durch suggestive Befragungen unwahrscheinlich ist (vgl. GREUEL, 2001).

Unter dem Aspekt der *Suggestion bei neutralen Beobachtungen* kann grundsätzlich in dem Befund, dass lediglich Gesehenes generell schlechter erinnert wird als eigene Handlungen bzw. Erlebnisse, eine besondere Anfälligkeit für interrogative Suggestionseinflüsse gesehen werden. Weiterhin wird davon ausgegangen, dass die Aussagegenauigkeit von Kindern dann reduziert ist, wenn sie Erlebtes verstandesmäßig nicht erfassen können. Zusammenfassend kann festgestellt werden, dass auch bei reinen Beobachtungsleistungen eine hohe Aussagegenauigkeit und Suggestionsresistenz selbst jüngerer Kinder unter neutralen Befragungsbedingungen angenommen werden kann, wobei auch hier wiederum eine höhere Leistung in Bezug auf zentral-aktionale Beobachtungsdetails feststellbar ist (vgl. GREUEL, 2001). Des weiteren zeigt sich diesbezüglich, dass die Wiederholung offener Anstoßfragen zu zuverlässigen Aussagergänzungen führt, wobei vor allem bei jüngeren Vorschulkindern Beschreibungen von Personen und Objekten durch Fragewiederholungen in besonderem Maß an Zuverlässigkeit verlieren (Cassel & Bjorklund, 1995; Dent, 1991; Memon & Vartoukian, 1996; zit. nach GREUEL, 2001).

Zeugen, die im Rahmen von Ermittlungen falschen oder irreführenden Informationen ausgesetzt sind, werden oftmals nachfolgende (follow-up) Fragen gestellt, um detaillierte Informationen über Personen, Dinge oder Geschehnisse zu erhalten. In diesem Zusammenhang ist auch auf die Untersuchung von DRIVDAHL & ZARAGOZA (2001) hinzuweisen, die zu dem Ergebnis kam, dass, wenn Zeugen bedrängt werden, sich ausgiebig und sorgfältig mit perzeptuellen Merkmalen (vor allem Örtlichkeiten, physikalische Bedingungen) suggerierter Ereignisse zu beschäftigen oder diese zu erinnern, erhöht dies die Wahrscheinlichkeit unzutreffender Gedächtnisinhalte oder Erinnerungen. Die auf diese Weise befragten Versuchspersonen behaupteten später eher, sich definitiv daran zu erinnern, die suggerierten Items (im Video) gesehen zu haben als Versuchspersonen, die zwar denselben Suggestionen ausgesetzt waren, jedoch nicht nach zusätzlichen Details befragt wurden. Auch die Fähigkeit für bildhaftes Vorstellungsvermögen sowie die Präferenz eines bildhaften kognitiven Stils stehen in Zusammenhang mit der Empfänglichkeit für das Entstehen von 'false memories' (vgl. DRIVDAHL & ZARAGOZA, 2001).

Nach LINDSAY (2002) ergeben nach den Ergebnissen der Suggestibilitäts-Studien bezüglich entwicklungspsychologischer Veränderungen inkonsistente Befunde. Ein interessanter Aspekt dabei ist die Tatsache, dass suggestive Details im 'free recall' von Erwachsenen häufiger berichtet wurden als von Vorschulkindern. Dies könnte nach LINDSAY (2002) unter anderem folgende Gründe haben:

- die Einzelheiten der Ereignisse waren nicht gut rememberbar, da sie nebensächliche Details des einige Tage zuvor präsentierten Versuchsmaterials waren und die Einzelheiten des eigentlichen Ereignisses somit bei der Abfrage eher nicht im Gedächtnis abrufbar waren;
- die übliche, normale Instruktion hielt die Versuchspersonen nicht davon ab, die nachträglichen Informationen trotzdem zu nutzen;
- Erwachsene erinnerten sich deshalb besser an die 'post-event-Suggestionen', weil diese hinsichtlich des Enkodierens und Behaltens verbaler Informationen höher entwickelte Fähigkeiten haben als etwa Vorschulkinder.

Insgesamt kommt nach aktuellem Forschungsstand situativen Aspekten des Suggestionprozesses eine große Bedeutung zu. Demnach geht man in kognitiven Erklärungsansätzen davon aus, dass die Übernahme nachträglicher Fehlinformationen eine Funktion der Stärke der Gedächtnisspuren darstellt. Eine Fehlinformation wird demnach dann übernommen, wenn die Gedächtnisspur für das tatsächliche Ereignis schwach, die Gedächtnisspur für die suggerierte Information hingegen relativ stark ist. Wiederholte oder kombinierte suggestive Einflussnahmen, also starke Fehlinformationsspuren, ziehen dabei größere Aussageveränderungen nach sich als einmalige. Ferner spricht hierfür, dass eine Fehlinformation, die kurz nach einem Ereignis erfolgt, weniger Effekt hat als die Präsentation zu einem späteren

Zeitpunkt, wenn die ursprüngliche Gedächtnisspur bereits verblasst ist und die Fehlinformation in keinem Konflikt mit einer (nicht mehr vorhandenen) eigenen Erinnerung steht. Suggestionseffekte sind demnach nur innerhalb eines spezifischen Erinnerungsintervalls zu erwarten (Bruck et al., Volbert & Dippold, 1998; vgl. VOLBERT, 1997; GREUEL, 2001).

Von Interesse sind in diesem Zusammenhang auch das *Ausmaß der Überlappung eines nachfolgenden Interviews mit dem beobachteten Ereignis* bezüglich des narrativen Inhalts, Struktur und umgebenden Kontexts. Insgesamt kann festgehalten werden, dass die genannte Überlappung in den drei Dimensionen zu keiner Erhöhung suggestiver Effekte führte. Struktur und Inhalt des Kontextes, in dem eine Falschinformation auftritt, deren Interaktion sowie verschiedene Erinnerungskontexte (Wiedererkennen vs. Erinnern) sind für das weitere Verständnis von Quellenverwechslungen und Suggestionen im Rahmen von Augenzeugensituationen entscheidend (vgl. MITCHELL & ZARAGOZA, 2001).

Bei der Wirksamkeit von Suggestionen wird neben kognitiven Faktoren auch *die Bedeutung sozialer Faktoren* betont. So geht die *Hypothese der Antworttendenzen* (McCloskey & Zaragoza, 1985) davon aus, dass Fehlinformationen auch dann übernommen werden, wenn eine richtige Erinnerung an das ursprüngliche Ereignis besteht. Dies geschieht entweder, weil die beeinflusste Person ihre eigene Erinnerung als weniger zuverlässig einschätzt als die von einem kompetenten Dritten (bei Kindern oft Erwachsene als Autoritätsperson; bei intellektuell eingeschränkten Menschen evtl. solche ohne diese Einschränkungen) oder weil sich die beeinflusste Person entsprechend den angenommenen Erwartungen verhält. Eine solche, gerade für Kinder spezifische Anfälligkeit resultiert möglicherweise daraus, dass es für Kinder üblich ist, Informationen von Erwachsenen in kommunikativen Prozessen zu erwerben (vgl. VOLBERT, 1997).

Nach Meinung des Verfassers dieser Arbeit kann bei lernbehinderten Jugendlichen, die sich oftmals in ähnlichen Situationen sehen ebenfalls von einer ähnlich gearteten spezifischen Anfälligkeit ausgegangen werden. Demnach erscheint es umso wichtiger, sich in kommunikativen Prozessen der Begutachtung von Zeugenaussagen nicht suggestiv und der Situation angemessen zu verhalten.

In Bezug auf falsche Suggestionen sollten generell unmittelbare von zeitverzögerten (längerfristigen) Effekten unterschieden werden. Von unmittelbaren Effekten spricht man, wenn suggestiv lenkende Befragungen zu schlechteren Erinnerungsleistungen führen als nicht lenkende Befragungen. Unmittelbare Effekte werden scheinbar durch zwei unterschiedliche Arten von sozialen Einflüssen vermittelt. Zum einen sind normative Einflüsse zu nennen, die dazu führen, dass der falschen Suggestion entgegen der eigenen Überzeugung gefolgt wird, um durch Zustimmung soziale Anerkennung zu gewinnen. Zum anderen informative Einflüsse, bei denen der Befragte die falschen Suggestionen irrtümlicherweise als valide Informationsquelle nutzt, um die eigenen Wissenslücken zu schließen. Überraschenderweise gibt es

nur sehr wenige empirische Befunde über natürliche Einflussquellen, wie z.B. den Einfluss anderer Personen, da im Bereich der Sozialpsychologie in einigen Untersuchungen zu konformem Verhalten bei einfachen Wahrnehmungsurteilen deutliche soziale Einflüsse nachgewiesen wurden. Solche Einflüsse müssten bei Erinnerungsaufgaben in noch stärkerem Maße wirksam sein als bei Wahrnehmungsurteilen, da das Vertrauen auf das eigene Gedächtnis das Vertrauen auf die eigenen Wahrnehmung übertreffen sollte (vgl. SCHWARZ, ROEBERS & SCHNEIDER, 2004). So konnten GABBERT et al. (2003) zeigen, dass falsche Informationen, die von einer natürlichen, sozialen Einflussquelle vermittelt werden, zu mehr Fehlern in der späteren Erinnerungsleistung führen als eine nicht soziale Falschinformation gleichen Inhalts (z.B. in Form eines Textes). Im Kontext von Untersuchungen zur Konformitätsentwicklung vor allem gegenüber Gleichaltrigen zeigte sich sowohl bei öffentlichen Urteilen in der Gruppensituation als auch bei privatem Urteil, bei dem Kinder Informationen über die angeblichen Urteile der anderen Teilnehmer oder Versuchspersonen erhielten, dass konformes Verhalten zum Jugendalter hin zunimmt (vgl. SCHWARZ et al. 2004). In ihrer eigenen Untersuchung beschäftigten sich SCHWARZ et al. mit Entwicklungsveränderungen bei sozialer Beeinflussung sowie deren Auswirkungen auf die Erinnerungsleistung. Dabei wurden 89 Kinder im Alter von 7 bis 8 bzw. 9 bis 10 Jahren nach einer Woche zum Inhalt eines Films befragt. Im ersten Interview wurde die Erinnerungsleistung über offene Fragen ermittelt. Anschließend wurde die Stärke des sozialen Einflusses von irreführenden und richtig lenkenden Suggestionen zusätzlich durch die Anwesenheit und das Antwortverhalten eines Verbündeten der Interviewer variiert. Am nächsten Tag wurde die suggestive Befragung ohne Verbündete wiederholt. Es zeigten sich deutliche Alterseffekte in der Fähigkeit sozialer Beeinflussung zu widerstehen. Während die 7 bis 8-Jährigen noch durch Suggestionen sowie zusätzlich durch Verbündete beeinflussbar waren, bestand bei 9 bis 10-Jährigen nur noch ein Einfluss der Verbündeten. Dies zeigte sich auch in altersspezifischen Unterschieden in den Nachwirkungen der Befragungssituation mit der Verbündeten auf die spätere Befragung ohne Verbündete (vgl. SCHWARZ et al., 2004).

Auch SHAW et al. (1997) weisen in ihrer Untersuchung auf den Einfluss der Information anderer Zeugen desselben Ereignisses (*co-witness information*) hin. Die Autoren untersuchten den kombinierten Effekt solch einer Information und suggestiver Befragung auf die Richtigkeit der Augenzeugenberichte hin. In drei Experimenten erhielten die Versuchspersonen jeweils falsche Informationen entweder durch den Interviewer (hinsichtlich dessen, was die anderen Zeugen bereits ausgesagt hätten) oder durch andere Zeugen selbst. In allen drei Experimenten gaben die Versuchspersonen eher die vorgegebene falsche Antwortalternative wieder bzw. schlossen sich dieser an. Dies galt nicht, wenn diese keine oder eine korrekte Information durch die 'co-witnesses' erhalten hatten. Dieser Effekt war auch 48

Stunden nach der ersten Befragung in einem weiteren Erinnerungstest feststellbar bzw. setzte sich fort (vgl. SHAW et al., 1997).

WRIGHT et al. (2001) konnten in ihrer Untersuchung nachweisen, dass die Erinnerung an präsentierte Szenen durch post-event Informationen weniger zugänglich wird.

Sowohl auf Seiten des Zeugen (Befragten) als auch des Befragenden sind feste Überzeugungen sowie die Vernachlässigung von Alternativhypothesen für das Verständnis suggestiver Prozesse von zentraler Bedeutung. Die Auswirkungen vorgefaßter Überzeugungen und Meinungen lassen sich aus der *Theorie der kognitiven Dissonanz* ableiten, wonach nach einmal getroffenen Entscheidungen der eigenen Überzeugung widersprechende (dissonante) Informationen einen Spannungszustand auslösen, der reduziert werden muss/soll. Dies geschieht unter anderem dadurch, dass bevorzugt solche Informationen gesucht werden, die die vorhandenen Überzeugungen und Erinnerungen bestätigen, während diejenigen, die ihnen widersprechen gemieden werden (*confirmation bias*) (vgl. KÖHNKEN, 1999B).

Hinsichtlich des Einflusses des Befragungsstils, der Interviewer-Konsistenz und der Persönlichkeit von Kindern auf die Suggestibilität und Quellenidentifikationsfehler veröffentlichten aktuell SHAPIRO & PURDY (2005) eine Untersuchung, in der Kinder mit („erzwungenen“) Konfabulationen und Falschinformationen konfrontiert wurden. Sogar nach einer Woche war die Suggestibilität in der Gruppe ‘erzwungene Konfabulation’ (FCG) höher als in der suggestiven Gruppe (SG). Quellenfehler im Interview bei Konfabulationen waren in der FCG höher, wenn die Fragen von derselben Person gestellt wurden; in der SG, wenn diese von einer anderen Person gestellt wurden. Beide Gruppen glaubten allerdings fälschlicherweise, sie hätten zuvor Kontroll-Items eines wahren Ereignisses diskutiert. Die Suggestibilität erwies sich als hoch, wenn einfallsreiche und distanzierte Kinder konfabulieren sollten. Kontroll-Items eines falschen Geschehens wurden von unterwürfigen Kindern fälschlicherweise dem Interview zugeschrieben und in der FCG von ideenreichen Kindern dem Video. Schüchterne Kinder in der FCG attribuierten Konfabulationen fälschlicherweise dem Video zu, wogegen ablenkbare Kinder in der SG Kontroll-Items des falschen Geschehens beiden Quellen falsch zuordneten (vgl. SHAPIRO & PURDY, 2005).

Somit interagieren vermutlich soziale und kognitive Faktoren bei der Produktion von Suggestionseffekten, wobei nach VOLBERT (1997) eine Differenzierung zwischen aktiver (als von der Person ausgehendem Einfluss) und passiver Suggestion (Empfänglichkeit für Suggestionen) hilfreich erscheint. Der Zustand des beeinflussbaren Individuums kann als spezifische Mangelsituation beschrieben werden, welche als Ziel der passiven Suggestion ausgeglichen werden muss. Dabei kann es sich um affektive Bedürfnisse (Mangel an Liebe, Vertrauen, Sicherheit, Selbstwertgefühl), kognitive Bedürfnisse (Mangel an Erinnerung, Wissen, logischem Denken, Verständnis) oder um strukturelle Bedürfnisse (ungenügende Klarheit der Situation) handeln (vgl. VOLBERT, 1997).

Demnach beeinflussen insgesamt wahrscheinlich eine Vielzahl von Faktoren den bewussten und unbewussten Gebrauch zusätzlicher Information außerhalb des eigentlichen Ereignisses sowohl im experimentellen Setting (eyewitness misinformation paradigm) als auch in realen Augenzeugensituationen (vgl. LINDSAY, 2002):

- Die (erinnerungswürdige) Besonderheit des Inhalts der Ereignisdetails und der Suggestionen
- Die Plausibilität der Suggestionen
- Das Ausmaß, zu dem Abfragebedingungen die Verlässlichkeit von Informationen außerhalb des Geschehens als Antwortquellen fördern bzw. verringern
- Die Memorabilität sowie der Nutzen quellenspezifischer Information bei der Erinnerung der ereignisspezifischen sowie suggestiven Einzelheiten
- Die Fähigkeit quellenspezifische Gedächtnisinformationen in Verbindung mit automatischen und bewusst kontrollierten Entscheidungsprozessen zu nutzen um die Quellen der Erinnerung zu identifizieren

GREUEL (2001) verweist zusammenfassend darauf, dass im Zusammenhang mit Suggestionseffekten die Handhabung von Begriffen wie 'false memory', 'implanted memory' oder 'misremembering' kritisch ist. Die derzeitigen empirischen Befunde zur interrogativen Verfälschbarkeit lassen demnach nur Rückschlüsse auf die Aussageperformanz zu, nicht aber auf die zugrunde liegende Kompetenz oder gar Gedächtnisrepräsentation. Somit sei es angemessener, Suggestibilität als Beeinflussung von Aussagen (und nicht der Erinnerung selbst) zu definieren (vgl. GREUEL, 2001).

Bei sachgemäßer, suggestionsfreier Befragung manifestieren sich Erinnerungsmängel jüngerer Kinder primär in *Auslassungsfehlern (omission errors)* und weniger in falschen *Übertragungsfehlern (commission errors)*. Mangelnde Genauigkeit in Erlebnisberichten jüngerer Kinder bedeutet also zunächst einmal und vor allem Unvollständigkeit und nicht Fehlerhaftigkeit. Im aussagepsychologischen Kontext kennt man die generelle menschliche Neigung zur fragmentarischen Wiedergabe von Wahrnehmungseignissen als sogenanntes *Inkadenzphänomen*. Ergänzend sei an dieser Stelle wiederum auf die von der Grundlagenforschung weitgehend unberücksichtigte motivationale Komponente des Erinnerns hingewiesen (vgl. GREUEL, 2001). Diese scheint jedoch gerade bei Kindern, Jugendlichen und auch Jugendlichen mit Lernbehinderungen von entscheidender Bedeutung in Bezug auf die Aussagepsychologie zu sein, was noch diskutiert werden wird.

1.2.1.3 sonstige Aspekte

Von besonderem Interesse innerhalb der Entwicklungspsychologie ist vermehrt die Zuverlässigkeit von Berichten von Kindern über autobiographische Ereignisse und die für die Förderung korrekter bzw. unkorrekter Schilderungen verantwortlichen Mechanismen. Im Falle der Rolle als Zeuge oder Opfer einer Straftat sind hier besonders zwei Aspekte von Bedeutung. Zum einen die Art und Weise der Aufdeckung oder Enthüllung traumatischer Erfahrungen durch Kinder, wobei hier vor allem kognitive, motivationale und emotionale Einflussfaktoren eine Rolle spielen. Zum anderen, die Frage, ob und inwieweit suggestive Befragungen sich in erfundenen Schilderungen nicht-erlebter (traumatischer) Erfahrungen widerspiegeln. BRUCK & CECI (2004) weisen diesbezüglich in ihrem Artikel auf *vier große Missverständnisse bzw. fälschliche Annahmen* hin, die leider Auswirkungen sowohl bis in die Gerichtssäle haben, sondern auch auf die nachfolgende Forschung. Diese können hier nur kurz angedeutet werden (vgl. BRUCK & CECI, 2004):

- (1) *Sexuell missbrauchte Kinder räumen aus Gründen von Scham, Schuld und Angst ungern und unfreiwillig den Missbrauch ein.* Nach eingehender methodischer Kritik an einigen Studien, die dies nach ihren zweifelhaften Befunden behaupten und dem erneuten Hinweis auf adäquate und fundierte Befragungstechniken (nicht suggestiv, offen), gilt nach Ansicht der Autoren: Wenn missbrauchte Kinder direkt danach gefragt werden, enthüllen sie auch den Missbrauch und berichten darüber.
- (2) *Suggestive Interviews lassen sich rein an der Anzahl von leitenden Fragen festmachen.* Vielmehr habe der so genannte 'interviewer bias' großen Einfluss auf die letztliche Qualität der Erinnerungsleistung. Voreingestellte Überzeugungen auf Seiten der Befrager und eine darauf gründende Art der Befragung mit ihren fatalen Konsequenzen sowie selektive Verstärkung im Sinne derjenigen Informationen, die der Interviewer „hören will“ bzw. die in dessen internalisierte Vorstellung des Geschehens passen, wirken sich – bei Einflussnahme in einem früheren Interview - lang anhaltend negativ auch in späteren, unbeeinflussten Befragungen auf die Richtigkeit der Aussage aus.
- (3) *Suggestibilität sei vor allem ein Problem von Vorschulkindern.* Trotz einer Vielzahl entsprechender Literatur, finden sich vor allem in neueren Untersuchungen ebenso Belege dafür, dass sich das Phänomen der Suggestibilität auch auf Kinder mittleren Alters übertragen lässt und dass diese unter bestimmten Umständen diesbezüglich sogar noch anfälliger sind.

- (4) *Mehrere wiederholte suggestive Befragungen seien notwendig, um eine Aussage zu verfälschen oder zu beeinflussen.* Die Autoren verneinen dies und betonen mit Verweis auf aktuelle Forschungsergebnisse, dass schon ein einziges suggestiv geführtes Interview mindestens denselben negativen, lang anhaltenden Verfälschungseffekt haben kann. Dabei sind bei mehreren Interviews der Zeitraum sowie die Gedächtnisstärke des ursprünglichen Ereignisses entscheidend.

Augenzeugen traumatischer Ereignisse sprechen oft typischerweise über diese aus unterschiedlichen Gründen. Das Interesse einer Untersuchung von MARSH et al. (2005) bestand in der Frage, ob qualitativ unterschiedliche Schilderungen zu Unterschieden in der späteren Erinnerung führen. Alle Versuchspersonen der Studie sahen eine Szene aus einem Film mit gewalttätigem Inhalt, wobei ein Drittel im Anschluss über ihre emotionalen Reaktionen auf den Film sprach (im Sinne einer Unterhaltung unter Freunden), ein Drittel beschrieb die Ereignisse des Films (in der Art einer polizeilichen Vernehmung) und ein weiteres Drittel stellte die Kontrollgruppe ohne besondere Aufgaben. Nach einem Delay wurden alle Versuchspersonen bezüglich ihrer Erinnerungen an die Filmszene befragt. Das Sprechen über die Gefühle führte zu einer besseren Erinnerung der eigenen Gefühle, jedoch auch zu Subjektivität und einem höheren Anteil entscheidender Fehler im freien Bericht. Geringere Unterschiede fanden sich bei Testaufgaben, die mehr Behaltenshinweise lieferten, was darauf hinweisen könnte, dass die Konsequenzen wiederholter Erzählungen für die Erinnerung größer sind, wenn die betreffenden Personen ihre eigenen Behaltensstrukturen generieren müssen (vgl. MARSH et al., 2005).

SEAMON et al. (2002) weisen in ihrer Studie außerdem darauf hin, dass sowohl korrekte wie auch falsche Erinnerungen abnehmen, je länger das Behaltensintervall ist. Die Beharrlichkeit falscher Erinnerungen zeigte sich beim Erinnern wie auch beim Wiedererkennen, wobei das Ausmaß für falsche Erinnerungen höher war als für richtige sowie leichter im 'recall' als in der 'recognition' zu beobachten war (vgl. SEAMON et al., 2002). FROST et al. (2002) betonen die hohe Beständigkeit falscher Erinnerungen über einen langen Zeitraum hinweg und eine entsprechende Tendenz bei Versuchspersonen, irreführende Informationen vermeintlich gesehen zu haben. Dieser Effekt erweist sich nach längerem Behaltensintervall (1 Woche) als sogar größer gegenüber kurz nach Enkodierung der Information (nach etwa 10 Minuten).

Im Bereich autobiographischer Erinnerungen sollten in Untersuchungen von Hyman & Billings (zit. nach DRIVDAHL & ZARAGOZA, 2001) Versuchspersonen sich Gedächtnisinhalte gänzlich erfundener Ereignisse ihrer Kindheit vorstellen bzw. diese eventuell kreieren. Im nicht überraschenden Ergebnis konnte festgestellt werden, dass die Empfänglichkeit für 'false memories' von vollständigen autobiographischen Ereignissen geringer war, als eine

falsche Erinnerung an einen Aspekt eines erlebten Ereignisses zu produzieren. Demgegenüber erscheinen die Formen von Quellenverwechslungsfehlern in der Studie von DRIVDAHL & ZARAGOZA (2001) alltäglicher und eher als Nebenprodukt normaler kognitiver Prozesse. Im Zusammenhang mit kognitiven Prozessen sind auch immer *bestimmte Erinnerungs- und Gedächtnisfehler* zu nennen, insbesondere alle Arten von Falschinformationseffekten sowie *‘retroaktive Interferenz-Effekte’*. Letztere treten dann auf, wenn Zeugen in Situationen sind oder Ereignisse erleben oder beobachten, die dem ursprünglich beobachteten Ereignis (Verbrechen, Straftat) ähnlich sind, verglichen mit unähnlichen Ereignissen und reduzieren die Erinnerung oder das Wiedererkennen des Verbrechens. CHANDLER et al. (2001) kamen zu dem Ergebnis, dass eine Gedächtnisspur für ein Ereignis nicht durch das Beobachten ähnlicher Ereignisse selbst verändert wird, sondern dass *‘postevents’* das Behalten stören. Versuchspersonen erinnerten in einem direkt anschließenden Erinnerungstest, Einzelheiten der ursprünglichen Geschichte weniger oft, wenn in einer nachfolgenden Geschichte bestimmte Dinge erwähnt wurden als wenn dies nicht der Fall war. Der Interferenz-Effekt trat dann auf, wenn die Versuchspersonen Details beider Geschichten erinnern sollten, jedoch nicht, wenn sie in erster Linie Items aus der ersten Story erinnern sollten. Ein eher allgemeiner Interferenz-Effekt konnte nach einem Tag festgestellt werden. Wurde die zweite Geschichte kurz vor der Abfrage präsentiert, wurden weniger Zielobjekte der ursprünglichen Geschichte erinnert als wenn beide sich einen Tag früher ereigneten. Die zweite Geschichte hatte somit nur dann störenden Einfluss auf die Erinnerung der ersten, wenn sie kontextuelle Behaltenshinweise betonte, die nicht mit der Gedächtnisspur der Zielobjekte zusammenpassen oder übereinstimmen (vgl. CHANDLER et al., 2001).

HANNIGAN & REINITZ (2001) weisen in ihrer Untersuchung auf zwei Arten von inferenzbasierten Fehlern, zum einen so genannte *‘schematic gap-filling’ errors*, zum anderen *‘causal-inference-based’ errors*. In drei Experimenten sahen die Versuchspersonen Dias von alltäglichen Routineaktivitäten, wie beispielsweise Einkaufen von Lebensmitteln. Neben dem zu erwartenden Effekt, dass beide Fehlerarten über die Zeit des Behaltensintervalls zunahmen, betonen die Autoren einen grundlegenden Unterschied von *‘schematic gap-filling’ errors* gegenüber *‘causal-inference-based’ errors*. Erstere scheinen mit „Wissensantworten“ assoziierbar und tendenziell auf einem Gefühl der Vertrautheit zu basieren ohne explizite Erinnerung. Letztere scheinen eher mit falscher episodischer Erinnerung einherzugehen. Nach Hannigan & Reinitz ergeben sich hieraus theoretische wie angewandte Implikationen. Möglicherweise verwechseln Menschen manchmal ihre schema-basierten und kausalen Inferenzen. Gerade im forensischen Kontext und Situationen im Gerichtssaal ist dies von Bedeutung, da mit zunehmendem Zeitintervall auch die Wahrscheinlichkeit von Inferenzfehlern steigt. Die Autoren nehmen an, dass Inferenzen als Einheiten im Gedächtnis repräsentiert sind, die über Handlungssequenzen hinweg wandern können, so dass eine Inferenz, die

in einem bestimmten Kontext auftritt, später in einem anderen Kontext eine falsche Erinnerung bezüglich der Geschehnisse verursacht. Alternativ dazu könnten Inferenzen auch stark mit einem spezifischen Kontext verbunden sein oder sogar die zugrunde liegende episodische Repräsentation einer spezifischen Ereignissequenz verändern (vgl. HANNIGAN & REINITZ, 2001).

In ihrer schon erwähnten Untersuchung von 1998 gingen GREENBERG et al. der Frage nach, unter welchen Bedingungen Ereignis-Schemata (Skripte) so genannte *'gap-filling-errors'* in Augenzeugenberichten eines Überfalls produzieren. Entsprechend der Hypothesen fand sich eine höhere Fehlerrate beim fälschlichen Wiedererkennen von zentralen gegenüber peripheren Handlungen. Je länger das Behaltensintervall, desto höher war auch diese Tendenz, Lücken mit erwartungsgemäßen Informationen zu füllen (GREENBERG et al., 1998).

Auch TUCKEY & BREWER (2003) untersuchten, auf welche Weise das Schema eines Verbrechens die verschiedenen Arten von Details beeinflusst und wie die Zeugen dies vor der ersten sowie über und zwischen mehreren, unterschiedlich zeitverzögerten Befragungen erinnert haben. Allgemein konnte festgestellt werden, dass schema-irrelevante Gedächtnisspuren nach dem ersten Interview schneller und in größerem Umfang zerfallen als schema-konsistente und schema-inkonsistente. Zudem wirkte sich eine Verzögerung einer ersten Befragung negativ auf die Gedächtnisleistung bei dieser ersten Befragung aus, führte jedoch auch zu weniger Verfall über die nachfolgenden Interviews hinweg. Hinsichtlich der Mehrdeutigkeit des (kriminellen) Reizmaterials, konnte gezeigt werden, dass Zeugen ihre verfügbaren Schemata nutzen, um mehrdeutige Informationen zu interpretieren und demnach mehr schema-konsistente Intrusionen, weniger korrekte Antworten produzierten sowie auch eher unzutreffende Gedächtnisinhalte schildern, die eine bewusste Erinnerung umfassen. Die Qualität der Enkodierung beeinflusst dabei die anfängliche Stärke der assoziativen Verbindungen. Die Stärke nimmt zwar über die Zeit hinweg ab, jedoch stellt jedes nachfolgende Behalten die Stärke der Verbindung teilweise wieder her. Einige episodische Knoten werden aufgrund des Ausmaßes der Verarbeitung oder der Anzahl der Wiederholungen zunächst besser und tiefer enkodiert als andere. Schema-inkonsistente Aspekte eines Verbrechens (welche den Vorstellungen eines schema-basierten Ereignisses widersprechen oder nicht erwartet werden) sind ungewöhnlich und kennzeichnend und sollten deshalb auf hohem Niveau verarbeitet werden mit dem Ergebnis von relativ starken Netzwerkverknüpfungen. Schemakonsistente Elemente werden wahrscheinlich auf einer mittleren Verarbeitungstiefe enkodiert, wohingegen sich schema-irrelevante Elemente aufgrund oberflächlicher Verarbeitung in eher schwachen Verknüpfungen manifestieren (vgl. TUCKEY & BREWER, 2003).

Im Rahmen ihrer Untersuchung weisen FORD et al. (2004) auf den oftmals beobachtbaren Befund hin, dass sich bei wiederholter Befragung über die Zeit hinweg die Gesamt-Quote

erinnerter Details erhöht (*Hypermnesie*). Solch eine Strategie fördert das Aufrechterhalten von Gedächtnisinhalten auf hohem Niveau. Allerdings können wiederholte Befragungen neben dem Aspekt der unsachgemäßen (suggestiven) Befragung auch andere negative Begleiterscheinungen mit sich bringen. Wenn Zeugen zu ausgewählten Aspekten einzelner Kategorien der beobachteten Informationen interviewt werden, erinnern sie später möglicherweise nicht-wiederholte Informationen derselben Kategorie schlechter als nicht-wiederholte Informationen einer anderen Kategorie. Dieser paradoxe Effekt ist als so genanntes *'retrieval-induced forgetting'* bekannt (vgl. auch Shaw et al., 1995; zit. nach LA ROOY et al., 2005). In diesem Zusammenhang ergeben sich aus der vorliegenden Arbeit aus dem angewandten Bereich daher eventuell auch wiederum Implikationen für die Grundlagenforschung, die noch zu diskutieren sind. Somit erweist es sich - unter anderem aufgrund der Ergebnisse der Untersuchung von FORD et al. – sowohl bezüglich Kinder als auch Erwachsener als kritisch für den Interviewer, einen umfassenden Rückblick oder eine vollständige Überprüfung des beobachteten Geschehens zu fördern (vgl. FORD et al., 2004). Wiederholte Erinnerungsversuche (z.B. in Form von immer wiederkehrenden Befragungen) können unter bestimmten Umständen nicht zu Hypermnesie, sondern auch zu so genannter *Reminiszenz* (Erinnerung bzw. Enthüllung neuer Information) führen. Dabei ist Hypermnesie abhängig von Reminiszenz, jedoch auch von der Erinnerung an vorher bereits erinnerte Information. Da wiederholte Erinnerungsversuche gewöhnlich auch Auslassungen zuvor erinnerter Informationen (also Vergessen) sowie auch Reminiszenzen beinhalten, sollte diese (für neue Informationen) für eine beobachtete Hypermnesie das Vergessen übertreffen. Zwar existieren diesbezüglich einige Untersuchungen bei Erwachsenen, jedoch nicht bei Kindern, Jugendlichen oder lernbehinderten Jugendlichen. So fanden beispielsweise Kern et al. (2002) bzw. auch Dunning & Stern (1992, beide zit. nach LA ROOY, 2005) ein vermehrtes Ausmaß an Hypermnesie bei Erwachsenen im Falle von emotional negativem und aufrüttelndem Stimulusmaterial gegenüber neutralem Versuchsmaterial. Möglicherweise führen offene Befragungstechniken beim Augenzeugengedächtnis von Kindern zu einer gesteigerten Erinnerung in Form von Reminiszenz und Hypermnesie. Daher mag teilweise auch die Kontroverse hinsichtlich wiederholter Befragungen im forensischen Kontext bzw. der realen Welt herrühren und dazu beitragen, dass entsprechende voneinander abweichende Struktur-, Quantitäts- und Qualitätsmerkmale von Augenzeugenaussagen der selben Person über dasselbe Ereignis für eher unglaublich erachtet werden (vgl. LA ROOY et al., 2005). HENKEL (2004) fand, dass Versuchspersonen über mehrere Erinnerungsversuche hinweg mehr Fehler in Form von Quellenverwechslungen unterliefen bei der Entscheidung, ob die zu erinnernden Stimuli als Bilder oder Worte präsentiert wurden. Dies galt insbesondere dann, wenn die Versuchspersonen sich nicht bewusst waren, dass sie später zu der Quelle der Erinnerung befragt werden.

Die schon mehrfach erwähnte Studie von HENRY & GUDJONSSON (2003) kam bei 11- und 12-Jährigen mit und ohne Lernbehinderung ebenfalls zum Ergebnis eines Anstiegs der freien Erinnerungsleistung über zwei Interviews hinweg, jedoch nicht hinsichtlich spezifischer geschlossener Fragen (vgl. HENRY & GUDJONSSON, 2003). Auch wenn Kinder keine Hypermnesie zeigen, erinnern sie möglicherweise neue Informationen über wiederholte Befragungen und Erinnerungsversuche hinweg. Im Fall einer gleich bleibenden oder über die Zeit abnehmenden Erinnerungsleistung folgt daraus nicht notwendigerweise, dass die gleiche Information einfach von Interview zu Interview übernommen wird. Reminiszenz (ohne Hypermnesie) konnte in einigen Studien beobachtet werden, die insgesamt zu dem Ergebnis kamen, dass neue Informationen in späteren Befragungen aufgedeckt und in die Schilderung mit aufgenommen und einbezogen werden kann. Dabei ist allerdings der Grad der Genauigkeit bzw. Korrektheit der Informationen zu beachten. Insgesamt kann festgestellt werden, dass die neue Information in der Regel weniger genau und richtig ist als diejenige, die bereits konsistent über verschiedene Interviews hinweg berichtet wurde (Salomon & Pipe, 1997, 2000; Peterson et al., 2001; zit. nach LA ROOY et al., 2005). In ihrer eigenen Untersuchung (LA ROOY et al., 2005) erwies sich Reminiszenz als zuverlässiger Indikator, obwohl die Anzahl neu erinnerter Items nach einem Delay geringer war als bei sofortiger Befragung. Hypermnesie konnte nur in einem Experiment nachgewiesen werden, als die Befragung gleich im Anschluss bzw. 24 Stunden nach dem Ereignis stattfand (gegenüber nach 6 Monaten in den anderen Experimenten). Die Gesamtzahl der innerhalb eines Interviews berichteten Fehler stieg bei Befragung nach 5 Minuten oder 24 Stunden in keinem der Experimente an, wogegen die kumulierte Erinnerung neuer Fehler anstieg (vgl. LA ROOY, 2005).

Von besonderem Interesse für die vorliegende Dissertation ist die Untersuchung von HENKEL (2004). In drei Experimenten wurde die Auswirkung von wiederholten und verlängerten Erinnerungsversuchen bezüglich der Ausmaße falscher Erinnerung eingeschätzt (*'paradoxical effects of repeated testing'*). Im Einzelnen sollte untersucht werden, ob wiederholte Erinnerungsversuche möglicherweise nicht nur die absolute Gesamtmenge an erinnerter Information erhöht, sondern eventuell gleichzeitig die Genauigkeit der Quellenbestimmung reduziert; zum zweiten sollte der Einfluss verschiedener Arten von Ähnlichkeit auf Quellenverwechslung durch wiederholte Befragung analysiert und weiterhin bestimmt werden, ob sich daraus ergebende falsche Quellenzuschreibungen durch sorgsam bedachte Beurteilung und Einschätzung während der Erinnerung vermindert werden können. Dazu sahen die Versuchspersonen (im Durchschnitt im Alter zwischen 20 und 23 Jahren) Bilder und stellten sich ebenso welche vor und vollzogen dann mehrere, wiederholte Erinnerungsversuche, bevor sie an einer Quellenidentifikationsaufgabe teilnahmen. Im Ergebnis ergaben sich folgende Befunde: Obwohl die Anzahl der erinnerten Items mit wiederholten Befragungen zunahm, zeigten sich daneben auch vermehrt Quellenfehler (d.h. vorgestellte Bilder

wurden als „gesehen“ eingestuft). Vorgestellte Items wurden dann eher fälschlicherweise als „gesehen“ eingestuft, wenn sich diese entweder physikalisch ähnlich waren oder semantisch-konzeptuell mit einem gesehenen Item in Beziehung standen. Wenn die Versuchspersonen wiederholt Items erinnerten ohne sich dabei die jeweilige dazugehörige Quelle zu überlegen, fanden sich nachfolgende Quellenfehler häufiger als wenn dies stattfand. Dieselben Effekte konnten sowohl für mehrere, einzelne, wiederholte aufeinander folgende Befragungen als auch für ein längeres Abfrageverfahren festgestellt werden. Die genannten Befunde decken sich mit der allgemeinen Einschätzung der möglichen gegensätzlichen Effekte wiederholter Befragungen auf die Gedächtnisleistung. Ein Anstieg an erinnerten Details geht demnach immer mit vermehrten Quellenverwechslungen dieser erinnerten Items einher. Zudem werden diese falschen Erinnerungen häufig vermeintlich zuverlässig als „gesehen“ eingeschätzt (vgl. HENKEL, 2004). Bei Erwachsenen waren nach Bornstein et al. (1998; zit. nach HENKEL, 2004) wiederholte Erinnerungsversuche von Einzelheiten eines Verbrechens zwar sowohl mit dem Erinnern mehr korrekter Informationen verbunden wie jedoch auch mit einer (geringen, aber doch deutlichen) Zunahme von Intrusionen. Solche Erinnerungsfehler könnten systematische Verzerrungen zum Ausdruck bringen, die Teil normaler kognitiver Prozesse sind. Somit sind diesbezüglich immer die Kosten gegenüber dem Nutzen wiederholter Befragungen in Form der Fehleranzahl gegenüber der Zahl der korrekten Details gegeneinander abzuwägen.

Neben den genannten Faktoren ist beim Wiedererkennen von Personen wie auch im freien Bericht über Ereignisse auch der Einfluss des Alters von Bedeutung (ROEBERS & SCHNEIDER, 2001B; WRIGHT & STROUD, 2002). WRIGHT & STROUD (2002) fanden dabei, dass ältere Personen besser und leichter ebenfalls ältere Beschuldigte/Verdächtige wieder erkennen als jüngere Verdächtige. Analoges gilt für jüngere Personen als Zeugen (vgl. WRIGHT & STROUD, 2002). Erwachsene erweisen sich hinsichtlich der korrekten Zurückweisung bei der Wiedererkennung von Gesichtern gegenüber Kindern als besser. Die freie Erinnerungsleistung verbessert sich während der Kindheit mit zunehmendem Alter. Die Konsistenz der verbalen Erinnerungsleistung und Fähigkeit zur Gesichtererkennung konnte bei ROEBERS & SCHNEIDER (2001B) nur für die Gruppe der 10-Jährigen festgestellt werden. (vgl. ROEBERS & SCHNEIDER, 2001B).

Mit Bezug auf ihre früheren Arbeiten mit Erwachsenen (Koriat & Goldsmith, 1994, 1996) weisen KORIAT et al. (2001) auf verschiedene wichtige Befunde hin: Demnach können Zeugen die Genauigkeit ihrer Aussage wesentlich erhöhen, wenn sie (a) effektiv in der Überwachung der Korrektheit ihrer Antworten sind, (b) frei in ihrer Entscheidung sind, ihre Schilderung entsprechend zu kontrollieren und (c) sie motiviert und angespornt werden, genaue Schilderungen abzugeben. Im Ergebnis ihrer Untersuchung kommen KORIAT et al. (2001) zu dem Schluss, dass sowohl 7- bis 9-jährige als auch 10- bis 12-jährige Kinder die

Genauigkeit ihrer Aussagen erhöhen können, wenn sie im freien Bericht falsche Antworten aussondern. Im Übrigen bestätigen die Befunde von 2001 im Wesentlichen die vorangegangenen.

Der Focus von ROEBERS & SCHNEIDER (2001A) lag auf der Untersuchung des Einflusses insbesondere der Persönlichkeitsmerkmale 'Intelligenz' und 'Schüchternheit' auf das Augenzeugengedächtnis von Kindern im Alter zwischen 6 und 10 Jahren. Insgesamt konnte festgestellt werden, dass Intelligenz einen signifikanten Einfluss auf die Leistung im freien Bericht (free recall) hatte, jedoch nicht auf die Suggestibilität. Schüchternheit hatte ebenso signifikanten Einfluss auf die Genauigkeit im 'cued recall' (auf Nachfrage bzw. mit Hinweisreizen). Während der Effekt der Intelligenz positiv mit der Erinnerung korrelierte, fand sich bezüglich der Schüchternheit ein negativer Zusammenhang. Schüchterne Kinder waren demnach bei der Beantwortung spezifischer Fragen weniger genau und zuverlässig als gleichaltrige, nicht schüchterne Kinder. Weiterhin weisen die Autoren darauf hin, dass mit der Persönlichkeit eines Menschen kongruente Informationen ein Selbst-Schema bzw. eine internes Arbeitsmodell aktivieren, was sich in besseren Erinnerungsleitungen kongruenter Ereignisse manifestiert. Somit haben bestimmte Persönlichkeitseigenschaften signifikanten Einfluss auf das Enkodieren und die Erinnerung komplexer Ereignisse (vgl. ROEBERS & SCHNEIDER, 2001A).

In dem gerade dargestellten Zusammenhang nennt und ergänzt ROBERTS (2003) neben den bekannten und untersuchten Einflüssen und Techniken zur Veränderung und Manipulation von Gedächtnisinhalten weitere natürlich auftretende Variablen (Faktoren im Bereich der Persönlichkeit), über die im experimentellen Setting wie auch in der forensischen Realität der Befragte keine oder nur wenig Kontrolle hat; insbesondere der Einfluss individueller Unterschiede hinsichtlich von Depression, (charakterlicher) Ängstlichkeit, Vorstellungsvermögen und auch der Interaktion dieser Faktoren unter Bedingungen von Stress und Belastung. Unter Stresseinfluss sind nach ROBERTS (2003) Personen mit lebendigem, bildhaftem Vorstellungsvermögen anfälliger für fehlerhafte Erinnerungen, ebenso wie auch weniger ängstliche Personen Erinnerungseinschätzungen eher einem falschen Gedächtnisinhalt zuschreiben.

Die Wissensbasis von Menschen wird von vielen Quellen beeinflusst, unter anderem auch von fiktionalen wie Filme oder Geschichten. MARSH et al. (2003) konnten zeigen, dass und wie Informationen aus fiktionalen Quellen von Personen gelernt und in das allgemeine Wissen über die Welt integriert werden. Diese Informationen werden im Gedächtnis sowohl mit verwandtem prä-experimentellem Wissen und der Quelle der Geschichte verbunden. Dies kann unter anderem mit einer gewissen Bereitschaft erklärt werden, Fakten aus einer Geschichte in einem allgemeinen Wissenstest gebrauchen und einbringen zu wollen.

Besonders kritisch im Rahmen der Forschung der Zeugenpsychologie sowie des sexuellen Missbrauchs bei Kindern ist die Fähigkeit der Kinder zwischen ähnlichen Vorfällen und spezifischen Details wiederholter Ereignisse differenzieren zu können, insbesondere nach einer zeitlichen Verzögerung. Gerade bei Kindern und intelligenzgeminderten Personen stellt dies ein tiefgründiges Problem dar, da deren Gedächtnis- und Erinnerungsprozesse nicht so gut entwickelt sind, wie die von Erwachsenen. Fehler treten hier in erster Linie in Form von internalen Intrusionen, also der Übertragung bzw. Integration von Details anderer Vorfälle auf bzw. in den aktuell zu erinnernden Vorfall. Je jünger die Kinder und je länger das Behaltensintervall, desto größer die Anzahl internaler Intrusionsfehler. Berichtete externe Intrusionen (falsche Details die in keinem der Vorfälle des Ereignisses vorkamen) sind von Kindern, die ein Ereignis wiederholt erfahren haben, weniger zu erwarten. Obwohl Alter und Behaltensintervall ebenfalls die Fähigkeit von Kindern beeinflusst, ein einzelnes Ereignis zu erinnern, ist der Einfluss der genannten Faktoren größer, wenn Kinder Vorfälle eines wiederholten, veränderlichen Geschehens erinnern sollen. Sowohl bei der Schilderung einzelner wie auch wiederholter Ereignisse legen die Befunde der Forschung nahe, dass eine Spezifizierung der Fragestellung einerseits mit einem Anstieg korrekter wie auch falscher Details einhergeht. Die Fehlerzahl steigt deshalb, da bei einer Abfrage im Falle nicht vorhandener oder nicht zugänglicher Informationen, das Kind raten muss bzw. eine Antwort generiert, die im Interesse des Befragenden liegt, dessen Vorstellungen entspricht oder einfach eine auf ähnliche Erfahrungen bezogene Antwort gibt, um nicht keine Antwort zu geben. Im freien Bericht ergeben sich dabei auch immer mehr Anhaltspunkte, die die einzelnen Vorfälle innerhalb der Ereignisse voneinander unterscheidbar machen (POWELL & THOMSON, 2003).

1.2.2 Aspekte der Glaubhaftigkeitsbegutachtung

Psychologische Sachverständige werden vorwiegend in Verfahren wegen vermuteter Straftaten gegen das sexuelle Selbstbestimmungsrecht damit beauftragt, die Glaubhaftigkeit von Zeugenaussagen zu begutachten. Dabei sieht der Bundesgerichtshof (BGH) im Falle kindlicher Zeugen eine Sachverständigenbegutachtung als erforderlich an, wenn im konkreten Einzelfall entwicklungspsychologische Einschränkungen gegeben sind (BGH NStZ 1981, 400). Bezüglich etwaiger Entwicklungsauffälligkeiten rekurriert die Justiz ebenfalls in erster Linie auf kognitiv-funktionale Einschränkungen, sei es, dass kognitiv-intellektuelle oder sozio-emotionale Entwicklungsdefizite eine Rolle spielen, sei es, dass es im Allgemeinen um Einschränkungen der intellektuellen Leistungsfähigkeit wie im Fall von Minderbegabungen und/oder geistigen Behinderungen geht. Somit wird also primär auf die Aussagetüchtigkeit kindlicher und jugendlicher Zeugen im Sinne der Wahrnehmungs-, Erinnerungs- und Aus-

drucksfähigkeit abgehoben (vgl. GREUEL, 2001). Nachstehende Zusammenfassung gibt einen Überblick über die in der forensischen Praxis als relevant erachteten Einschränkungen bzw. Beurteilungsschwierigkeiten im Falle kindlicher und jugendlicher Zeugen.

Nach GREUEL (2001) lassen sich folgende Anlässe für die Begutachtung kindlicher und jugendlicher Zeugen zusammenfassend nennen (aus GREUEL, 2001, S. 9):

- Niedriges Lebensalter
- Langes Erinnerungsintervall
- Entwicklungsdefizite
- Einschränkungen der kognitiven Leistungsfähigkeit
- Gefahr sekundärer Traumatisierung
- Gefahr suggestiver Aussagebeeinflussungen

Im Sinne eines Opfer- bzw. Zeugenschutzes wird von legislativer Seite die möglichst frühe Hinzuziehung psychologischer Sachverständiger empfohlen, um Mehrfachvernehmungen, die damit eventuell viktimisierenden Auswirkungen und/oder eventuellen Beeinträchtigungen oder Verfälschungen der Aussagen von Kindern und (intellektuell beeinträchtigten) Jugendlichen zu verhindern. Auf die Problematik suggestiver Aussagebeeinflussungen und Einflüsse durch nachträgliche und/oder falsche Informationen von verschiedenen Seiten gerade der speziellen Personengruppe der Lern- oder geistig Behinderten wird im Verlauf der weiteren Darstellung noch näher eingegangen. Der Fokus im Rahmen der „Symbiose“ von Justiz und (Aussage-)Psychologie verschiebt sich somit immer mehr weg von motivationalen Erörterungen hin zu einer stärkeren Berücksichtigung der Aussagetüchtigkeit (vgl. GREUEL, 2001).

Anlässe für die Begutachtung erwachsener Zeugen sind in erster Linie dem psychiatrischen und klinischen Bereich zuzuordnen. Dabei handelt es sich auch auf dem Hintergrund rechtlicher Vorgaben vor allem um solche Zeugen, bei denen aufgrund klinisch bedeutsamer Auffälligkeiten im Erleben und Verhalten und/oder intellektueller Leistungsdefizite (z.B. Lern- oder geistige Behinderung) Zweifel an der Aussagetüchtigkeit bzw. der Suggestibilitätsneigung entstehen können und um Fragen, inwieweit der Zeuge zur Erstattung einer gerichtsverwertbaren Aussage imstande ist.

Somit ergeben sich vor allem folgende Anlässe für die Begutachtung erwachsener Zeugen (GREUEL, 2001, S. 12):

- Intellektuelle Leistungsdefizite
- Suchtproblematik
- Persönlichkeitsstörungen
- Psychosomatische Störungen

- Psychiatrische Störungen
- Neurologische Störungen

Wenn aussagepsychologische Begutachtung unter anderem durch die Widerlegung des Vorurteils von der „unbrauchbaren“ Kinderaussage oder des Anzweifeln der „Daseinsberechtigung“ von Zeugenaussagen intellektuell beeinträchtigter Menschen zur Stärkung der Position von Zeugen und deren Rehabilitation beitragen kann - was *ein* Anliegen dieser Arbeit ist – bzw. was durch qualifizierte Begutachtung und aus empirischer Sicht von erlebnisfundierten Aussagen erwartet werden kann und muss, ist die Forderung nach systematischer Etablierung aussagepsychologischer Glaubhaftigkeitsbegutachtung generell logisch konsequent und wünschenswert (vgl. auch GREUEL, 2001).

1.2.2.1 Die aussagepsychologische Konstrukt-Trias

Nach GREUEL (2001) ist der eher umgangssprachliche Ausdruck der 'Glaubhaftigkeit' aus aussagepsychologischer Sicht als mehrdimensionales Konstrukt zu verstehen. Es werden die drei übergeordnete psychologische Konstrukte angesprochen:

Aussagetüchtigkeit – Aussagequalität – Aussagevalidität

Auf der Ebene der *Aussagetüchtigkeit* stellt sich die Frage, ob und in welchem Ausmaß eine Person über entsprechende kognitiv-funktionale Leistungsvoraussetzungen verfügt. Insbesondere sind hier Prozesse der verbalen Rekonstruktion eines subjektiven Erlebnisses, also primär Aspekte der Wahrnehmungs-, Erinnerungs- und Ausdrucksfähigkeit immer in Bezug auf die konkreten relevanten Rahmenbedingungen angesprochen. Weiterhin wird das Konstrukt der Aussagetüchtigkeit wiederum selbst maßgeblich von folgenden psychologischen Dimensionen bestimmt (vgl. GREUEL, 2001):

- Wahrnehmung
- (Episodisches) Gedächtnis
- Sprachliches Ausdrucksvermögen
- Sprachverständnis
- Konzentrationsvermögen
- Wirklichkeitsverständnis
- Suggestibilität

Wie noch detailliert dargestellt werden wird, sind einige der genannten Faktoren gerade bei Menschen mit kognitiven Einschränkungen nicht vollständig ausgebildet bzw. gestört und deshalb im Rahmen dieser Arbeit von besonderem Interesse.

Allerdings ist hierbei abzugrenzen, dass unter der Zeugentüchtigkeit kein situationsübergreifendes, personales Merkmal verstanden wird. Generell gehen GREUEL et al. (1998) davon aus, dass rein die Orientierung an einem Intelligenzniveau unzureichend erscheint bei der Beantwortung der Frage, ob ein Zeuge in der Lage ist, einen Sachverhalt zuverlässig wahrzunehmen und wiederzugeben, da „es kein festgelegtes Intelligenzniveau gibt, das für die Eignung als Zeuge erforderlich ist“ (S.88). Auch Personen mit begrenzten intellektuellen Fähigkeiten seien zu genauer Beobachtung und wirklichkeitsgetreuem Behalten im Gedächtnis in der Lage. In diesem Kontext zeigt sich, dass die Beurteilung der Aussagetüchtigkeit also immer individuell erfolgen muss und keinesfalls am grundsätzlichen Vorliegen einer Intelligenzminderung festgemacht werden darf.

Weiterhin betont MÜLLER-LUCKMANN (1980), dass die Kompetenzfrage durchaus situationsspezifisch zu beurteilen ist. So können minderbegabte Zeugen beispielsweise Schwierigkeiten haben, einen als Augenzeuge erlebten, komplexen Verkehrsunfall zu schildern, andererseits aber durchaus in der Lage sein, Aussagen über ein selbst erlebtes Delikt gegen die sexuelle Selbstbestimmung zu machen.

Neben der Feststellung der Aussagetüchtigkeit liegt es im Aufgabenbereich des psychologischen Sachverständigen, das Gericht auf Einschränkungen der speziellen Aussagetüchtigkeit aufmerksam zu machen, auf spezielle Techniken bei der Zeugenbefragung zu verweisen und die Gewichtung einzelner Aussageelemente darzulegen.

Unter dem Aspekt der *qualitativen Ausprägung einer Aussage* gilt es zu klären, ob die Aussage solche Merkmale aufweist, die in erlebnisbasierten Schilderungen zu erwarten sind, in frei erfundenen und/oder von Dritten induzierten Aussagen oder konstruierten, intentionalen Falschaussagen hingegen fehlen, diese also voneinander zu differenzieren sind.

Grundlage dafür sind die in der traditionellen Forschung als „Glaubhaftigkeitskriterien“ oder „Realkennzeichen“ bezeichneten Qualitätsmerkmale, die auf Basis der so genannten ‘UNDEUTSCH-HYPOTHESE’ (Undeutsch, 1967) entwickelt wurden:

Erlebnisbegründete Aussagen unterscheiden sich von intentionalen Falschaussagen sowohl in qualitativer als auch in quantitativer Hinsicht bezüglich ihres globalen inhaltlichen Erscheinungsbildes voneinander, und zwar derart, dass erlebnisbezogene Aussagen höhere Ausprägungsgrade der inhaltlichen Realkennzeichen aufweisen als Falschaussagen.

Eine in diesem Zusammenhang der Distinktheit erlebnisbasierter Aussagen auftretende Frage zielt auf die möglicherweise zugrunde liegenden Prozesse. Zwar heben sich einige Untersuchungen und Konzeptionen von einer reinen experimentellen Validierung der singulären Qualitätsmerkmale ab, indem sie auf spezifische Charakteristika abheben, wie kognitionspsychologische Aspekte der Wirklichkeitswahrnehmung (STADLER, 1997) oder Realitätsüberwachung (SPORER & KÜPPER, 1995; SPORER, 1997), jedoch geht es auch hier primär um Fragen des 'wie, sich Repräsentationen unterschiedlicher Wirklichkeitsstufen unterscheiden', denn um das 'warum, sich erlebnisfundierte Aussagen von konstruierten Aussage ohne Erlebnishintergrund' unterscheiden. Insofern kann die vorliegende Arbeit als Versuch angesehen werden, zumindest beide Aspekte simultan und an derselben speziellen Population zu beleuchten und eventuell teilweise einen Integrationsversuch darzustellen. Eine Antwort auf die eben gestellte Frage könnte nach GREUEL (2001) lauten: Autobiographische Erlebnisse werden auf eine spezifische Art und Weise enkodiert und gedächtnismäßig repräsentiert, wogegen Erlebnisse, die nicht auf das individuelle Selbst bezogen sind und damit nicht als autonoetisch erfahren werden, können auch nicht ebenso repräsentiert und letztendlich reproduziert werden (vg. GREUEL, 2001).

Abschließend geht es um die Frage, ob eine qualitativ hochwertige und insofern auf eine Erlebnisbasis verweisende Aussage unter Berücksichtigung individueller (interner) und kontextueller (externer) Rahmenbedingungen auch als zuverlässig beurteilt werden kann. Nach GREUEL (2001) lassen sich drei Gruppen von Faktoren unterscheiden, die die *Aussagevalidität (Aussagezuverlässigkeit)* beeinflussen können:

- Psychologische Besonderheiten der Aussageperson
(Leistungsbesonderheiten, Persönlichkeitsstörungen)
- Motivationale Bedingungen der Aussage
(Aggravationstendenzen, Wahrnehmungsselektion)
- Kommunikative Bedingungen der Aussagegenese und –entwicklung
(Suggestive Beeinflussungen, aber auch z. T. nicht-intentionale Falschinformationseffekte)

Dies deckt sich zum Teil ebenfalls mit den von KÖHNKEN (2001) zusammengestellten wesentlichen Elementen einer aussagepsychologischen Untersuchung zum Zwecke der Glaubwürdigkeitsbeurteilung:

- (1) die Untersuchung relevanter Aspekte der Persönlichkeit des Zeugen;
- (2) die Analyse der Entstehungs- und Entwicklungsgeschichte der Aussage;
- (3) die kriterienorientierte Analyse des Aussageinhalts anhand der erläuterten Realkennzeichen;

- (4) die systematische Überprüfung verschiedener Hypothesen, die die Aussage mit den darin festgestellten Qualitätsmerkmalen (Realkennzeichen) erklären könnten.

Innerhalb der Persönlichkeitsdiagnostik gilt es neben der Analyse aussagerelevanter Kompetenzen des Zeugen zur Erhebung der erwähnten wichtigen individuellen Vergleichsstandards, im Rahmen einer so genannten *Persönlichkeits- und Kompetenzanalyse* noch ein zweiten Aspekt zu erheben. Hier geht es allgemein formuliert um die Feststellung bereichsspezifischer Kompetenzen und Erfahrungen des Zeugen, wie dessen allgemeine und sprachliche intellektuelle Leistungsfähigkeit, seine kreativen Fähigkeiten und im Fall von Sexualdelikten der Einschätzung diesbezüglicher Kenntnisse und Erfahrungen. Als multimethodale Analyse- und Messinstrumente dieser persönlichen Kompetenzen eines Zeugen dienen die allgemeinen Methoden der Diagnostik, wie Befragung, Beobachtung, (standardisierte) Tests und Fragebögen, biographische Analysen und Ähnliches (vgl. STELLER, 1999). Zu beachten sind in diesem Zusammenhang Phänomene und/oder Zeugen, die Auffälligkeiten oder Beeinträchtigungen bezüglich der Wahrnehmungsorganisation oder gar Persönlichkeitsstörungen aufweisen, da sich hier aufgrund spezifischer Akzentuierungstendenzen erhebliche Einschränkungen in Bezug auf die Validität der Aussage ergeben. Dabei ist für die Glaubhaftigkeitsbegutachtung stets im Einzelfall zu prüfen, wie gravierend die Beeinträchtigung der Realitätsorientierung ist und auch dem intraindividuellen Vergleichsstandard bei solchen Verdachtsfällen ein besonderer Stellenwert beizumessen (vgl. GREUEL et al., 1998). Insgesamt betrachtet kann die Kompetenzanalyse somit nach STELLER (1999) einerseits als Bezugspunkt für die Qualitätsanalyse fungieren, andererseits aber auch zur Beurteilung der Aussagetüchtigkeit, also der prinzipiellen Fähigkeit des Zeugen zur Wahrnehmung, Speicherung und Reproduktion von Sachverhalten sowie zur Realitätskontrolle dienen.

Neben der Persönlichkeits- und Kompetenzanalyse muss nach STELLER & VOLBERT (1999) eine *Fehler- und Motivationsanalyse* quasi als Zwischenstufe zwischen Personen- und Aussageanalyse eingefügt werden. Hieraus wird die direkte Verknüpfung der Verwertbarkeit von Ergebnissen der Qualitätsanalyse mit der Angemessenheit von Befragungen und deren Rahmenbedingungen deutlich. Ein Teil der Fehlerquellenanalyse stellt die Motivationsanalyse dar, die gezielt nach motivationalen Tendenzen sucht, die sich verfälschend auf die Aussage ausgewirkt haben könnten. Dabei muss jedoch angemerkt werden, dass eine Handlung oft nicht nur durch *ein* Motiv, sondern durch das Zusammenwirken mehrerer Motive und Anregungsbedingungen in einer konkreten Situation motiviert ist, über deren Bedeutung oder Veranlassung die Betroffenen selbst oft nicht vollständig Auskunft geben können. Die Motivationsanalyse ist stets als ein Prozess nachträglicher Erklärung einer

einzelnen Handlung anzusehen. Dazu ist es allerdings in einem interpretativen Prozess erforderlich, die Verschränkung von Motivgrundlagen und äußeren Anregungsbedingungen sowie (wenn möglich) deren subjektive Bedeutung für den Handelnden in der jeweiligen Handlungssituation und zum Zeitpunkt der Aussage zu rekonstruieren, die dann letztendlich zu der fraglichen Handlung geführt haben (vgl. GREUEL et al., 1998). Es erscheint in diesem Zusammenhang sinnvoll, vor allem die Motivation der Erstaussage zu untersuchen, da bei eventuell weiteren Befragungen und Wiederholungen der Aussage zusätzliche Motive von Bedeutung sein können bzw. sich entwickeln. GREUEL et al. (1998) verweist dabei auf eine von ARNTZEN (1993) aufgestellte Reihe von potentiell wichtigen Informationen, wie z.B. besondere Entwicklungen und plötzliche Veränderungen in den sozialen Beziehungen des Zeugen, dessen äußere und psychische Situation zum Zeitpunkt der Erstaussage, aus denen oftmals Hinweise auf die damalige motivationale Situation des Zeugen ersichtlich werden (vgl. GREUEL et al., 1998). Wie schon erwähnt handelt es sich bei der Motivationsanalyse um einen rückblickenden Interpretations- und Rekonstruktionsprozess, für den allerdings keine Richtigkeitskriterien existieren, d.h. verschiedene Gutachter oder Personen rekonstruieren, interpretieren und gewichten die fragliche Motivation zum Aussagezeitpunkt unterschiedlich, wonach eine Beurteilung einer Aussage allein aufgrund der Motivationsanalyse wohl kaum ausreichend, jedoch recht unsicher wäre. Somit kommt also den Schlussfolgerungen aus der Motivationsanalyse im Vergleich zu den Ergebnissen der Aussageanalyse ein geringerer Stellenwert in der Glaubhaftigkeitsbegutachtung zu. Dennoch ist nach übereinstimmender Auffassung die Einbeziehung der Motivationsanalyse als weiteres unabhängiges Verfahren in aussagepsychologische Gutachten unbedingt erforderlich, da sie die Interpretationsbasis für Glaubhaftigkeitseinschätzungen in Kombination mit Kompetenz- und Qualitätsanalyse trotz ihrer Unspezifität erweitern kann (vgl. STELLER, 1999).

Ein letzter Bestandteil im Rahmen der Aussagevalidität stellt die *Analyse der Entstehungs- und Entwicklungsgeschichte der Aussage* dar. Hier sind vor allem Fragen zur Klärung möglicher suggestiver Beeinflussungen des Zeugen von zentraler Bedeutung. Es muss geklärt werden, wann und aus welchem Anlass zum ersten Mal welcher Person gegenüber Angaben über die inkriminierten oder fraglichen Geschehnisse gemacht wurden und welcher Art die Reaktionen anderer Personen darauf waren. Des Weiteren ist festzustellen, ob und in welcher Art und Umfang danach noch weitere Befragungen des Zeugen stattgefunden haben, was sich oftmals aus eventuell vorhandenen Ermittlungsakten nachvollziehen lässt. Ergeben sich aus diesen Fakten und Entwicklungen deutliche Hinweise auf eine erhebliche suggestive Beeinflussung, ist die kriterienorientierte Aussageanalyse als Beurteilungsinstrument zwischen wahren und erfundenen Schilderungen zu differenzieren, nicht mehr in vollem Ausmaß geeignet (vgl. KÖHNKEN, 2001). Während es also bei der Analyse der Entwick-

lungsgeschichte der Aussage vor allem darum geht zu untersuchen, ob es von der Erstbekundung bis zur aktuellen Aussage systematische Veränderungen gab und ob diese möglicherweise in Zusammenhang mit bestimmten Ereignissen (wie z.B. Gesprächen oder Befragungen) aufgetreten sind, ergibt sich des weiteren ein gewisser Überschneidungsbereich mit der Konstanzanalyse, die neben der Inhaltsanalyse ein weiteres wesentliches methodisches Element der Aussageanalyse ist und als Bestandteil der Qualitätsanalyse verstanden wird. Diese und deren Befunde sind dann umso mehr von Bedeutung, wenn der zeitliche Abstand zwischen dem tatsächlichen Ereignis und dessen Schilderung sehr groß ist oder wenn der Zeuge während des geschilderten Erlebnisses großem Stress oder emotionaler Belastung ausgesetzt war (vgl. KÖHNKEN, 2001; STELLER, 1999).

Somit gilt festzuhalten, dass das Urteil der Glaubhaftigkeit einer Aussage nie auf der Basis einer Konstruktebene zu fällen ist, sondern vielmehr auf Grundlage einer integrativen Betrachtung der Befunde aller drei Konstruktebenen.

1.2.2.2 Integrative Kriteriologie nach Steller und Köhnken (1989)

Auf der Grundlage der oben erläuterten Ausgangshypothese ('UNDEUTSCH-HYPOTHESE') und verschiedenen anderen, teilweise unterschiedlichen Kriteriensystemen und Merkmalslisten (z.B. Undeutsch (1967), später Arntzen (1993), Littmann & Szewczyk (1983) sowie Trankell (1971)) stellten STELLER & KÖHNKEN (1989) eine integrative Kriteriologie aus 19 Realkennzeichen bzw. Glaubhaftigkeitsmerkmalen zusammen (vgl. KÖHNKEN, 2001). Diese Kategorisierung wurde zunächst in englischer Sprache als „Criteria-Based Content Analysis“ (CBCA; Kriterienorientierte Aussageanalyse) beschrieben (vgl. STELLER; WELLERSHAUS, WOLF, 1992). Diese liegt inzwischen nahezu allen neueren empirischen Studien zur inhaltsorientierten Glaubwürdigkeitsbeurteilung zugrunde und soll im folgenden Kapitel dargestellt werden.

Tabelle 1-3: Realkennzeichen nach STELLER & KÖHNKEN (1989)

<p>Allgemeine Merkmale</p> <ol style="list-style-type: none"> 1. Logische Konsistenz / Widerspruchsfreiheit 2. Unstrukturierte Darstellung / ungeordnete Reproduktionsweise 3. Quantitativer Detailreichtum <p>Spezielle Inhalte</p> <ol style="list-style-type: none"> 4. Raum-zeitliche Verknüpfungen (= Kontextuelle Einbettung) 5. Interaktionsschilderungen 6. Wiedergabe von Gesprächen (= wörtliche Rede) 7. Schilderung unvorhergesehener Komplikationen im Handlungsverlauf <p>Inhaltliche Besonderheiten</p> <ol style="list-style-type: none"> 8. Schilderung ausgefallener / ungewöhnlicher Einzelheiten 9. Schilderung nebensächlicher Einzelheiten 10. Phänomengemäße Schilderungen unverständener Handlungselemente 11. Indirekt handlungsbezogene Schilderungen / externe Assoziationen 12. Schilderung eigener psychischer Vorgänge 13. Schilderung psychischer Vorgänge des Täters / beim Beschuldigten <p>Motivationsbezogene Inhalte</p> <ol style="list-style-type: none"> 14. Spontane Verbesserungen der eigenen Aussage 15. Zugeben von Erinnerungslücken 16. Einwände gegen die (Zweifel an der) Richtigkeit der eigenen Aussage 17. Selbstbelastungen 18. Entlastung des Beschuldigten <p>Deliktspezifische Inhalte</p> <ol style="list-style-type: none"> 19. Beschreibungen deliktspezifischer Elemente / Merkmale
--

Obige Tabelle zeigt die 19 *inhaltlichen Realkennzeichen der Kriterienorientierten Aussageanalyse* in der Fassung von STELLER & KÖHNKEN (1989). Dabei handelt es sich um Merkmale *einer* Aussage. Realkennzeichen, die sich aus dem Vergleich wiederholter Aussagen (über denselben Sachverhalt) erschließen lassen (Konstanz, Widersprüche, Auslassungen und Ergänzungen), sind hier somit nicht enthalten, wobei diese Beschränkung systematische Gründe hat und keinerlei Wertung beinhaltet (vgl. STELLER, WELLERSHAUS, WOLF, 1992). Die Realkennzeichen insgesamt basieren auf der Annahme, dass sie in einer erfundenen Aussage nur selten vorkommen, weil ein Zeuge nicht in der Lage ist, eine Aussage mit den in den Realkennzeichen beschriebenen Qualitäten ohne eigene Erlebnisgrundlage zu erfinden (vgl. KÖHNKEN, 2001).

GREUEL (2001) weist allerdings darauf hin, dass die auf der Grundlage der qualitativen Aussageanalyse und auf Basis der Undeutsch-Hypothese entwickelten Qualitätsmerkmale in ihren Termini (Realkennzeichen, Glaubhaftigkeitsmerkmale oder Glaubhaftigkeitskriterien) kritisch und problematisch zu sehen sind (auch GREUEL et al., 1998). Auf diese aus ver-

schiedenen Gründen berechtigte Kritik kann hier nicht näher eingegangen werden, sei dem interessierten Leser jedoch zum Nachlesen empfohlen. GREUEL schlägt vor, demnach treffender von *Qualitätsmerkmalen erlebnisfundierter Aussagen* zu sprechen und dies auch in weiteren Forschungsarbeiten zumindest inhaltlich zu berücksichtigen. In der vorliegenden Arbeit sollen allerdings die Original-Begrifflichkeiten von STELLER & KÖHNKEN beibehalten werden.

Allgemeine Merkmale

Als Teil der Qualitätsanalyse umfasst die merkmalsorientierte Inhaltsanalyse fünf Kategorien. In der ersten Kategorie, den *allgemeinen Merkmalen* der Aussage, werden die ersten drei Realkennzeichen zusammengefasst, für deren Analyse die gesamte Aussage zu untersuchen ist und die sich auf deren globales Erscheinungsbild beziehen. Diese können ohne Rückgriff auf Einzelheiten des Aussageinhalts beurteilt werden. Dabei werden in der Literatur insbesondere die Merkmale 'Detaillierungsgrad', 'Anschaulichkeit', 'Strukturgleichheit' und 'Logische Konsistenz' (siehe dazu Tabelle 1 - 4) unter diesem Aspekt diskutiert und ihnen besonderes Gewicht bei der Beurteilung der Aussage zugewiesen (vgl. GREUEL et al., 1998, S. 92).

'**Logische Konsistenz**' bedeutet, dass eine Aussage frei von inneren Widersprüchen ist. Innere Stimmigkeit, logische Folgerichtigkeit und Widerspruchsfreiheit sind ohne Zweifel Grundvoraussetzungen, um eine Zeugenaussage als erlebnisbezogen zu qualifizieren (GREUEL et al., 1998, S. 96). Allerdings darf 'Logische Konsistenz' im Sinne dieses Realkennzeichens nicht automatisch mit vordergründiger Plausibilität gleichgesetzt werden. Unerfahrene Beobachter könnten hier manche Verhaltensweisen (z.B. von Missbrauchsopfern) als unplausibel bewerten und somit fälschlicherweise die Unglaubwürdigkeit der Aussage daraus ableiten (vgl. KÖHNKEN, 2001). Dieses Merkmal ist dann erfüllt, wenn sich die Einzelelemente der Gesamtaussage in einen schlüssigen, logisch folgerichtigen Zusammenhang bringen lassen, ohne dass sich dabei eben innerhalb der Aussage Widersprüche oder logische Brüche ergeben. In hoher Ausprägung liegt dieses Merkmal vor, wenn es sich um eine sehr umfangreiche, komplexe Aussage handelt. Wieder wird davon ausgegangen, dass es einem normal begabten falsch aussagenden Zeugen kaum gelingen kann, die einzelnen Bestandteile seiner Aussage jeweils so präsent zu haben, dass er sie im Rahmen einer Befragungssituation so aufeinander abstimmen kann, dass sich keine aussageimmanenten Widersprüche ergeben. Der damit verbundene kognitive Aufwand ist natürlich umso höher, je umfangreicher der Aussageinhalt ist (vgl. GREUEL et al., 1998).

Das Realkennzeichen **‘Unstrukturierte Darstellung’** begründet sich in der Beobachtung, dass lediglich erfundene Schilderungen in den meisten Fällen chronologisch vorgetragen werden. Dieses Kriterium gilt dann als erfüllt, wenn ein Zeuge seine Angaben ungeordnet, sprunghaft und unter wiederholtem Durchbrechen der äußeren Handlungschronologie macht. Die Belegkraft dieses Merkmals resultiert daraus, dass es einem falsch aussagenden Zeugen kaum möglich ist, auf einen chronologischen Aufbau und Abruf seiner Aussage zu verzichten und seine Aussage sprunghaft und unstrukturiert darzustellen (vgl. KÖHNKEN, 2001), während ein erlebnisbezogener Zeuge problemlos im zeitlichen Ablauf springen kann und sich oftmals die einzelnen Aussageelemente erst bei der Gesamtbetrachtung der Aussage in ein Gesamtbild integrieren lassen. Für das Auftreten dieses Qualitätsmerkmals ist jedoch ein größerer Aussagumfang erforderlich. Eine besonders hohe Qualifizierung erlangt dieses Merkmal insbesondere dann, wenn zum einen so genannte inverse Verlaufsstrukturen im Sinne rückwärts aufgerollter Interaktionsschilderungen in der Aussage vorhanden sind, zum anderen sich eine Vielzahl von über die Gesamtaussage verstreuten Einzelangaben und „Aussagefetzen“ in einen schlüssigen Gesamtzusammenhang integrieren lassen und sich die inhaltliche Struktur der Aussage oft erst am Ende der Befragung durch die Gesamtbetrachtung der Aussage erschließt (vgl. GREUEL et al., 1998). Nach Arntzen (1970, 1993) handelt es sich hierbei – im Falle von Kindern – um entwicklungspsychologisch bedingte Aussageeigentümlichkeiten, wonach sich bei Zeugen mit starker Tendenz zu unstrukturierter Aussageweise aufgrund von Defiziten bzw. nicht hinreichend entwickelten Fähigkeiten im Bereich des diskursiven Denkens hohe Ausprägungen im Sinne einer geringen Fähigkeit zur Steuerbarkeit von Erinnerungseinfällen ergeben (vgl. GREUEL et al., 1998, S. 101 f.). Nach GREUEL et al. (1998) ist die Auftretenswahrscheinlichkeit dieses Qualitätsmerkmals allerdings in starkem Maße von den kommunikativen Randbedingungen der Befragung abhängig. Der Zeuge muss seine Aussage im freien Bericht getätigt haben, um im nötigen Rahmen überhaupt unstrukturiert aussagen zu können. Bei Verstößen gegen die logische Konsistenz wird eine unstrukturierte Darstellung jedoch nicht im Sinne eines Realkennzeichens gewertet (vgl. STELLER, 1999).

Des weiteren besteht uneingeschränkter Konsens darüber, dass sich erlebnisgestützte Aussagen durch eine **Vielfalt von Detailangaben** (Merkmal des **‘Quantitativen Detailreichtums’**) auszeichnen. Dies ergibt sich aus dem Umstand, dass es für die meisten Zeugen bereits unter dem Aspekt der Informationsverarbeitungs- und Gedächtniskapazität unmöglich ist, eine Falschaussage mit vielen Details zu erfinden und zu handhaben. Das Merkmal des Detailreichtums ist dann erfüllt, wenn der Zeuge nicht nur vielfältige Details schildert, die im Sinne von Orts-, Handlungs- und Personenbeschreibungen auf das inkriminierte Geschehen bezogen sind, sondern zudem Details aus dem näheren Beziehungs- und Handlungsumfeld

mit einbezieht. Zudem dürfen sich keine wesentlichen Brüche ergeben zum einen zwischen dem Detaillierungsgrad von fallneutralen und fallbezogenen Schilderungen, zum anderen zwischen der Schilderung des fraglichen Kerngeschehens und den Angaben zu eher peripheren Kontextbedingungen. Dieses Merkmal hängt jedoch stark von Persönlichkeitsfaktoren wie dem Alter des Zeugen, dessen Sprachvermögen etc. ab. Nach KÖHNKEN (2001) basiert dieses Merkmal auf der Annahme, dass es für einen Zeugen ohne eigenen Erlebnishintergrund sehr schwierig ist, eine komplexe Aussage mit vielen Details zu erfinden und in sich stimmig zu produzieren. Was die operationale Definition dessen, was im Einzelfall als Detail zu werten sei, finden sich in der Literatur unterschiedliche Positionen, wobei sich nach der Mehrzahl der Ansichten der Detaillierungsgrad aus der Anzahl sämtlicher, in einer Aussage enthaltener Detailangaben, unabhängig von ihren qualitativen Besonderheiten ergibt. Dies steht im Einklang mit den bereits erwähnten hohen kognitiven Anforderungen an den Zeugen, die sich durch eine hohe Detailquantität ergeben. Zwar kann man wegen der Abhängigkeit dieses Merkmals von der Komplexität des entsprechenden Sachverhalts, der Länge des fraglichen Erinnerungsintervalls und dem habituellen Ausdrucksstil des Zeugen keine normierenden (absoluten) Häufigkeitsangaben machen, jedoch wird als Detailangabe jeder Aussagebestandteil gewertet, der der beschriebenen Situation oder Handlung ein neues Moment hinzufügt (vgl. GREUEL et al., 1998).

Spezielle Inhalte

In zwei weiteren Kategorien ('Spezielle Inhalte' und 'Inhaltliche Besonderheiten') sind solche Merkmale zusammengefasst, die sich auf einzelne Inhalte einer Aussage beziehen und bei denen nicht die bloße Quantität von geschilderten Details, sondern deren qualitative Ausprägung für die Analyse von Bedeutung ist. In der *Kategorie 'Spezielle Inhalte'* sind die folgenden vier Merkmale zusammengefasst: 'Kontextuelle Einbettung', 'Beschreibung von Interaktionen', 'Wiedergabe von Gesprächen' und 'Schilderung von Komplikationen im Handlungsverlauf'.

Das Realkennzeichen '**raum-zeitliche Verknüpfungen (= kontextuelle Einbettung)**' bezieht sich darauf und ist erfüllt, wenn in der Aussage vielfältige spezifische Verflechtungen des inkriminierten Geschehens mit veränderlichen situativen (örtlichen, zeitlichen, handlungsbezogenen und/oder biographischen) Umständen und Ausgangssituationen aus dem individuellen Umfeld des Zeugen und des Beschuldigten enthalten sind und aus denen heraus sich fragliche Tathandlungen entwickeln. Falschaussagende Zeugen müssten derartige Verflechtungen mit in Frage kommenden situativen Umständen nicht nur retrospektiv vornehmen, sondern sich auch über die Notwendigkeit derartiger Verankerungen im Klaren

sein. Nach der forensischen Erfahrung wird von falsch aussagenden Zeugen diese, ein hohes Maß an gedanklicher Logik und Vorstellungsfähigkeit erfordernde, intellektuelle Leistung weder überzeugend erbracht noch das Erfordernis einer derartigen Einbettung in Erwägung gezogen (vgl. Arntzen, 1993). Außerdem können Aussagen mittels dieser Kontexteinbettungen stets objektiv validiert und überprüft werden, was ein intentional falsch aussagender Zeuge ja in der Regel zu vermeiden bemüht ist (vgl. GREUEL et al., 1998). Je komplexer diese Verknüpfungen sind, desto schwieriger dürfte es sein, diese ohne eigene Erlebnisgrundlage zu (re-)produzieren (vgl. KÖHNKEN, 2001). Dieses Merkmal ist auch dann von herausragender Bedeutung, wenn die Möglichkeit einer Übertragung auf den Beschuldigten aufgrund von Erfahrungen mit anderen Personen zu prüfen ist (d.h. die Annahme, dass die beschriebenen Handlungen zwar tatsächlich erlebt wurden, aber mit einer anderen als der jetzt beschuldigten Person). Nach KÖHNKEN (2001) sind bei dieser Konstellation viele der übrigen Realkennzeichen möglicherweise nicht mehr trennscharf, da die Annahme einer vollständig erfundenen Aussage ohne Erlebnishintergrund hier nicht zutrifft.

Die Konstruktion und **'Beschreibung von** wechselseitig aufeinander bezogenen **Interaktionen'** im Sinne dieses Realkennzeichens ist bei Falschaussagen schwierig zu bewerkstelligen. Hierzu müsste der Zeuge psychologisch nachvollziehbare Sequenzen von aufeinander bezogenen Handlungen zwischen ihm und dem Beschuldigten im Sinne von Reaktion und Gegenreaktion beschreiben, die veranschaulichen, wie sich diese Verhaltensweisen gegenseitig bedingt haben. Auch dies ist bei Falschaussagen schwierig zu bewerkstelligen (vgl. KÖHNKEN, 2001). Die höchste Merkmalsausprägung und Aussagekraft haben dabei solche Schilderungen, die regelrechte miteinander verflochtene Reaktionsketten aufweisen (Interdependenz); je komplexer diese Interaktionsketten, desto stärker auch die Ausprägung des Realkennzeichens, besonders diejenigen, die sich auf das inkriminierte Tatgeschehen beziehen, allerdings zählen hierzu auch prä- und postdeliktische Interaktionssequenzen. Wenn Einzelhandlungen lediglich beschrieben oder ohne nähere Konkretisierung einfach nur benannt werden, ist nach GREUEL et al. (1998) dieses Merkmal nicht erfüllt (vgl. GREUEL et al., 1998).

Die **'Wiedergabe von Gesprächen'** zwischen dem Zeugen und dem Beschuldigten mit Bezug zum inkriminierten Geschehen stellt als Realkennzeichen quasi eine spezifische Art der Interaktionsschilderung auf verbaler Ebene dar und ist ihr strukturell ähnlich. Analog zur Handlungsebene muss auch hier zum einen ein gegenseitiger Wechsel von Gesprächsequenzen im Sinne von Rede und Gegenrede bzw. Frage und Antwort feststellbar sein. Darüber hinaus muss auch erkennbar sein, welche Inhalte von welcher Person geäußert

wurden, wobei diese Dialogschilderungen in direkter oder indirekter Rede erfolgen können (vgl. KÖHNKEN, 2001). Wenn dabei Schilderungen auftreten, die explizit die differentiellen (adversativen) Standpunkte der beiden Dialogpartner darstellen, so muss dem Realkennzeichen eine besonders hohe Qualifizierung beigemessen werden. Dies gilt ebenso für vom aussagenden Zeugen vorgebrachte, unverständene oder missverständene Äußerungen. Auch hier gilt, dass für die Wiedergabe derartiger Gesprächssequenzen ein besonderes Maß an logischer Kombinationsfähigkeit und der Fähigkeit zum Perspektivenwechsel und der Rollenübernahme notwendig wäre, um diese Inhalte in eine Falschaussage einflechten und ohne jeden Erlebnisbezug schlüssig produzieren zu können. Dies liegt aber außerhalb des Planungshorizontes eines Zeugen bzw. wird gar nicht erst in Erwägung gezogen (vgl. GREUEL et al., 1998).

Nach allgemeinem und ungeschränktem Konsens besteht ein weiterer starker Hinweiswert auf den Erlebnisbezug einer Aussage, wenn in dieser **‘Schilderungen von Komplikationen’**, von durch äußere Störeinflüsse bewirkte, missglückten oder vorzeitig abgebrochenen Handlungen, unerwartet auftretenden Hindernissen oder enttäuschten Erwartungen vorkommen. Für einen intentional falsch aussagenden Zeugen liegt es geradezu außerhalb seines Planungshorizonts, solche Details in seine Aussage zu implantieren. Da dies die Reproduktion einer Aussage erheblich erschwert, wird diese Möglichkeit in der Regel gar nicht erst bedacht und somit vermieden (vgl. GREUEL et al., 1998; KÖHNKEN, 2001). Nach GREUEL et al. (1998) lässt sich die geringe Auftretenswahrscheinlichkeit von negativen Komplikationen in intentionalen Falschaussagen zum einen motivations-, zum anderen auch kognitionspsychologisch erklären. Der „Lügner“ hat keinerlei Grund, gerade vom intendierten Belastungsschwerpunkt wegführende Handlungsverläufe oder zusätzlich komplizierende Handlungsaspekte in seine Aussage einzubauen. Auf der anderen Seite würde dies nur noch zusätzliche Anforderungen an dessen kognitive Leistungsfähigkeit stellen und das ‘handling’ seiner Falschaussage unnötigerweise erschweren (vgl. GREUEL et al., 1998).

Dieses Realkennzeichen erfasst neben Komplikationen auf der Handlungsebene auch solche auf der kognitiven Ebene, wie z.B. die angesprochenen enttäuschten Erwartungen oder gescheiterte Entwürfe von Handlungsplänen, wobei hier nach GREUEL et al. (1998) allerdings konkrete Handlungsansätze, die aus entsprechenden Überlegungen resultieren, geschildert werden müssen, damit das Merkmal erfüllt ist. Antizipierte Komplikationen sowie pauschale Benennungen von Komplikationen (wie z.B. „Ich habe versucht, mich zu wehren, aber es ging nicht“) zählen hierzu nicht (vgl. ARNTZEN, 1993).

Inhaltliche Besonderheiten

In einer weiteren Kategorie von *‘Inhaltlichen Besonderheiten’* werden folgende sechs Merkmale zusammengefasst: *‘Schilderung ausgefallener/ungewöhnlicher Einzelheiten’*, *‘Schilderung nebensächlicher Einzelheiten’*, *‘Phänomengemäße Schilderungen unverständener Handlungselemente’*, *‘Indirekt handlungsbezogene Schilderungen/externe Assoziationen’*, *‘Schilderung eigener psychischer Vorgänge’* und *‘Schilderung psychischer Vorgänge des Täters/beim Beschuldigten’*.

Die *‘Schilderung origineller, ausgefallener Einzelheiten’* zählt nach GREUEL et al. (1998) in der einschlägigen Fachliteratur zu den belegkräftigsten Hinweisen auf den Erlebnishintergrund einer Zeugenaussage. Darunter versteht man solche Details, die in irgendeiner Weise ausgefallen sind oder sich durch ihre geringe Auftretenswahrscheinlichkeit auszeichnen und deshalb nicht zu erwarten wären, weil sie unter anderem über schematisches Alltagswissen hinausgehen. Dabei dürfen diese jedoch im gegebenen, persönlichen Kontext und auf dem Hintergrund des intellektuellen Leistungspotential des Zeugen nicht unrealistisch sein und müssen somit immer an diesem relativiert werden. Die geringe Auftretenswahrscheinlichkeit kann hierbei zum einen in der Originalität des Details selbst begründet sein, zum anderen sich – z.B. in Kinderaussagen – aus den speziellen und kreativen Fähigkeiten des aussagenden Zeugen ergeben (vgl. KÖHNKEN, 2001). Nach GREUEL et al. (1998) ist „in solchen Fällen im Rahmen der Exploration abzuklären, ob derartige Detailangaben eine kindliche Metapher für ein Wahrnehmungserlebnis oder ein Phantasieprodukt darstellen“ (GREUEL et al., 1998, S. 119).

Bei *‘nebensächlichen oder überflüssigen Details’* handelt es sich nach KÖHNKEN (2001) um Einzelheiten, die mit dem inkriminierten Tatgeschehen selbst nichts zu tun haben (für einen sachlogischen Zusammenhang und dessen Verständnis also nicht notwendig sind) und für das relevante Kerngeschehen der Aussage meist unnötig sind, aber von der aussagenden Person im Zusammenhang mit dem Handlungsablauf in einer Vielzahl genannt werden. In erlebnisbezogenen Aussagen haben diese Details jedoch aufgrund des persönlichen Erlebens und eines persönlichen Relevanzsystems sowie der subjektiven Sicht des Zeugen durchaus eine gewisse, wichtige Bedeutung (vgl. GREUEL et al., 1998).

Auch hier geht man wieder davon aus, dass falsch aussagende Zeugen in der Regel nicht in Betracht ziehen, eine Fülle für die Beschuldigung *‘irrelevanter’* Details mit einzubeziehen und ihre Aussage so mit *‘kognitivem Ballast’* zu überfrachten. Außerdem erschwert dies den Überblick über die an sich schon konstruierte Falschaussage und deren Handhabung für den Zeugen (vgl. GREUEL et al., 1998).

Als Negativabgrenzung dieses als allein im quantitativen Sinn über die Gesamtaussage vorhandenen Merkmals, stellt GREUEL et al. (1998) fest, dass dieses Realkennzeichen nicht erfüllt ist, wenn der Bericht des Zeugen lediglich Abschweifungen enthält, die keinerlei Zusammenhang mit dem inkriminierten Geschehen aufweisen.

Eine **phänomengemäße Schilderungen unverstandener Handlungselemente** als weiteres Realkennzeichen liegt in Aussagen vor, die sich dadurch auszeichnen, dass ein Geschehen, Ereignis oder auch Interaktionsverläufe „naiv“, d.h. allein aufgrund seiner äußeren Anmutung konkret geschildert werden, ohne dass ein Zeuge deren Bedeutung richtig erfasst hat, nicht einordnen oder sich psychologisch nicht erklären kann. Jedoch ist der oder die Aussagende von deren Wahrnehmung überzeugt. Derartige Schilderungen finden sich hauptsächlich in Kinderaussagen über sexuelle Sachverhalte; sie können aber auch bei Jugendlichen oder Erwachsenen auftreten, wenn aus den Aussagen quasi die in der Vergangenheit noch vorhandene kindliche Wahrnehmungsebene deutlich wird oder diese ein Geschehen nur unvollständig wahrgenommen haben (vgl. GREUEL et al., 1998). KÖHNKEN (2001) bemerkt weiter, dass Kinder gelegentlich Wahrnehmungen, für die sie kein Schema besitzen im Wege der Assimilation (nach Piaget) einem schon vorhandenen Schema zuordnen. Dies wäre z.B. der Fall, wenn eine Ejakulation als „Spucken“ bezeichnet würde.

Indirekt handlungsbezogene Schilderungen als Realkennzeichen sind Schilderungen von Erlebnissen, die dem inkriminierten Geschehen ähnlich sind, sich aber zu einem anderen Zeitpunkt und mit anderen Personen ereignet haben (vgl. KÖHNKEN, 2001). Das Merkmal ist dann erfüllt, wenn der Zeuge über Gesprächsinhalte mit dem Beschuldigten im Kontext der fraglichen Tathandlungen berichtet. Diese Schilderungen stellen eine besondere Komplikationsform dar, da hier eine Verknüpfung der auf das Tatgeschehen bezogenen Gesprächsinhalte mit anderenorts gemachten, ähnlich gelagerten Erlebnissen, Erfahrungen oder Beziehungen des Zeugen vorgenommen wird. Solche – auch als **Inhaltliche Verschachtelungen** (Arntzen, 1993) bezeichnete – Schilderungen können dabei nicht nur zur Bestätigung des Erlebnishintergrundes einer Aussage herangezogen werden, sondern auch zur Abgrenzung des zur Debatte stehenden Vorfalls von anderen Situationen oder Personen (vgl. GREUEL et al., 1998).

GREUEL et al. (1998) betonen, dass davon als Negativabgrenzung jedoch Inhalte zu unterscheiden sind, in denen Zeugen lediglich Gespräche mit unbeteiligten Dritten über die fraglichen Ereignisse wiedergeben. Dann gilt das Merkmal als nicht erfüllt (vgl. GREUEL et al., 1998).

Als ein weiterer Hinweis auf den Erlebnisbezug einer Aussage kann gelten, wenn in einer Aussage **'Beschreibungen eigener psychischer Vorgänge'** in differenzierter Weise zu finden sind. Dies können Schilderungen über emotionale Prozesse, körperliche (physiologische) Empfindungen und/oder Reaktionen oder gedankliche Vorgänge (Reflexionen) sein, die mit dem Kerngeschehen zusammenhängen. Eine besondere Qualifizierung erfährt dieses Merkmal dann, wenn z.B. physiologische Vorgänge (wie Zittern, Herzrasen oder ähnliches) detailliert beschrieben werden oder Prozesshaftes im Sinne von Affektverläufen und Erlebnisentwicklungen (z.B. das Schildern des Erlebens des Errötens) geschildert werden. Wiederum gilt hier, dass es sehr hohe Anforderungen an die psychologische Abstraktions- und Differenzierungsfähigkeit eines Zeugen stellen würde, derartige Erlebnisweisen schlüssig in eine Falschaussage zu integrieren. Spontan vorgebrachten Angaben über eigenpsychisches Erleben oder auch die Schilderung komplexer affektiver Prozesse kommt dabei besonders große Bedeutung zu (vgl. Littman & Szewczyk, 1989 bzw. Arntzen, 1993; zit. nach GREUEL et al., 1998).

Auch hier lässt sich nach GREUEL et al. (1998) folgende Negativabgrenzung vornehmen: Eine ledigliche Benennung von einfachen und ohnehin nahe liegenden Gefühlsreaktionen (Angst, Schmerzen, Scham, Unwohlsein) ohne entsprechende qualitativ differenzierte Beschreibung hat für sich genommen keinen diagnostischen Wert, da diese auch leicht aufgrund von allgemeinem Schemawissen konstruiert und in eine Falschaussage integriert werden kann (vgl. GREUEL et al., 1998).

In Analogie zu dem vorstehendem Realkennzeichen beinhaltet auch die **'Beschreibung psychischer Vorgänge beim Beschuldigten'** prinzipiell die gleichen Phänomene. Auch hier werden Angaben zu emotionalen Reaktionen und/oder körperlichen/sensorischen Vorgängen auf Seiten des Beschuldigten als Indikatoren für den Erlebnisbezug der Aussage angesehen, wobei jedoch bisher noch keine bestimmten Anforderungen an die Differenziertheit solcher Angaben formuliert wurden. DAHLE & WOLF (1997) nahmen jedoch in einer Weiterentwicklung des Kriterienkatalogs von Steller & Köhnken (1989) eine inhaltliche Differenzierung dieses Aussagemerkmals vor. Hierbei subsumierten sie die Schilderung von sowohl äußerlich beobachtbaren als auch nicht beobachtbaren, aber indirekt erschlossenen gefühlsbezogenen motorischen und physiologischen Abläufen sowie die Schilderung von Gefühlen und Gedanken beteiligter Dritter unter diese Merkmalskategorie.

Für einen intentional falsch aussagenden Zeugen würde es nicht nur eine komplexe Phantasie- und Differenzierungsleistung darstellen, solche nicht erlebnisgestützten Beschreibungen psychischer Zustände Dritter zu konstruieren, sondern er müsste auch noch eine gewisse Fähigkeit zur Rollenübernahme haben oder entwickeln, wovon in der Regel – besonders bei Kindern – nicht auszugehen ist. Für diese stellt es sich besonders schwierig dar, sich in einen

andere Person hineinzusetzen, geschweige denn deren Gefühlslage oder Erleben zu schildern. Deshalb kann man diesem Merkmal aus psychologischer Sicht eine hohe Belegkraft zubilligen, noch mehr, wenn sich diese Angaben auf Prozesse und Affektentwicklungen beziehen (vgl. GREUEL et al., 1998).

Nach KÖHNKEN (2001) läge eine solche Schilderung im Sinne eines Realkennzeichens dann vor, wenn z.B. aus der Beobachtung von Erröten, Schwitzen, schneller Atmung etc. auf eine bei dem Beschuldigten vorhandene Erregung oder Aggression geschlossen wird.

Wiederum erfüllen auch hier dem allgemeinen Schemawissen entsprechende Beschreibungen einfacher und nahe liegender Gefühlsreaktionen noch nicht den Stand eines begutachtungsrelevanten Qualitätsmerkmals eines Erlebnishintergrunds.

Motivationsbezogene Inhalte

Die bisher dargestellten Realkennzeichen sind der kognitiven Komponente der `Undeutsch-Hypothese` zuzuordnen. Dahingegen bezieht sich eine vierte Gruppe von Merkmalen in der Kategorie `Motivationsbezogene Inhalte` auf die Selbstpräsentation des Aussagenden und ist somit der „impression-management-Komponente“ der `Undeutsch-Hypothese` zuzuordnen. Nach LEMPP, SCHÜTZE, KÖHNKEN (1999) vermeidet die Bezugnahme auf die Theorie des `Impression Management` (Tedeschi & Norman, 1985) mögliche Missverständnisse, da bei der Beurteilung dieser Realkennzeichen keine potentielle Motivation zur Falschbelastung untersucht wird. Finden sich dennoch derartige – scheinbar ungünstige – Äußerungen in der Aussage, kann man daraus schließen, dass die Erzeugung eines glaubwürdigen Eindrucks kein dominantes Motiv beim Zeugen war. Nach STELLER (1999) handelt es sich hier um inhaltliche Realkennzeichen, ohne dass direkt im Hinblick auf die Aussagemotivation exploriert wird. Dominierend ist also hier der motivationale Aspekt und es stellt sich die Frage, wie wahrscheinlich es ist, dass ein falsch aussagender Zeuge Inhalte der beschriebenen Art in seine Aussage aufnehmen würde (vgl. STELLER, 1999). Man geht hier von der oben stehenden und schon erläuterten Grundannahme einer unterschiedlichen Motivation von bewusst täuschenden und erlebnisbezogen aussagenden Zeugen aus. Unterstellt man ein auf einer gewissen Simplifizierung basierendes Modell der Gegenüberstellung des „zweckrationalen Lügners“ einerseits und dem objektiv berichtenden Zeugen andererseits, kann man lediglich die dargestellten subjektiven und als verhaltenswirksam unterstellten impliziten Glaubwürdigkeitskonzepte folgern. Alles in allem kommt der Literatur nach den motivationsbezogenen Aussagemerkmalen ein geringerer Hinweiswert auf den möglichen Erlebnisbezug einer Aussage zu, als im Vergleich dazu denjenigen, die sich durch kognitive Befunde und Gesetzmäßigkeiten begründen lassen (vgl. GREUEL et al., 1998).

In der einschlägigen Literatur besteht nach Ansicht der meisten Autoren ein weiterer hoher Beleg für den Erlebnisbezug einer Aussage in den **‘spontane Verbesserungen der eigenen Aussage’**. Darunter werden nach KÖHNKEN (2001) solche inhaltlichen (nicht grammatikalischen) Korrekturen verstanden, die ohne äußere Veranlassung wie z.B. das Vorhalten von Widersprüchen oder auf Nachfrage, erfolgen. Das Merkmal gilt als erfüllt, wenn diese Verbesserungen spontan von Seiten des Zeugen im Rahmen der Aussage oder auf dem Hintergrund einer erinnerungskritischen Haltung hervorgebracht werden, wobei ‘spontan’ nicht bedeutet, dass diese unmittelbar nach einer Detailschilderung erfolgen müssen. Wenn im Zuge der Schilderung anderer Handlungsabschnitte verschiedene Einzelheiten assoziiert werden, die vorher nicht oder anders erinnert wurde, so können die Korrekturen auch mit größerer zeitlicher Verzögerung erfolgen.

Dahinter steht die Annahme, dass solche Merkmalsausprägungen lediglich in Aussagen mit Erlebnishintergrund vorkommen, weil die entsprechenden Zeugen in der Regel an einer möglichst genauen Wiedergabe des Geschehens interessiert sind. Bei intentional falsch aussagenden Zeugen ist hingegen gerade ein Bemühen um die Vermeidung derartiger Korrekturen und Lücken erkennbar, da diese bestrebt sind, einen möglichst glaubwürdigen und erinnerungssicheren (vgl. Arntzen, 1993; Steller & Köhnken, 1989; zit. nach GREUEL et al., 1998) Eindruck zu hinterlassen.

Nach GREUEL et al. (1998) können sich diesbezüglich bei (mehreren) vorausgegangenen Aussagen manchmal Abgrenzungsprobleme zwischen Korrekturen und Inkonanzen ergeben.

Das Realkennzeichen **‘Zugeben von Erinnerungslücken’** geht von der schon mehrmals angeführten Hypothese aus, dass ein intentional falsch aussagender Zeuge auf jeden Fall darum bemüht sein wird, keine Lücken in seiner Darstellung erkennen zu lassen und somit dem Befrager keinen Hinweiswert auf eventuelle Aussagemängel liefern will. Ein Zeuge, der jedoch aufgrund eines tatsächlich sich ereigneten Geschehens berichtet und somit um Ehrlichkeit bemüht ist, kann also dagegen ohne Furcht vor Glaubwürdigkeitsverlust auch Erinnerungslücken und –ungenauigkeiten zugeben, da dies völlig natürliche Vorgänge sind (vgl. KÖHNKEN, 2001).

Jedoch ist nach GREUEL et al. (1998) der Hinweiswert dieses Kriteriums für den wahrscheinlichen Erlebnisbezug einer Aussage als eher gering einzustufen, da man auch bei intentional falsch aussagenden Zeugen erwarten kann, dass diese Erinnerungslücken zugeben, wenn sie z.B. auf eventuelle Nachfragen keine stimmigen Antworten vorbringen können.

Bezüglich der **‘Einwände gegen die Richtigkeit oder Glaubhaftigkeit der eigenen Aussage’** kann analog zu oben stehendem Realkennzeichen und auf dem Hintergrund des „zweckrationalen Lügners“ wiederum davon ausgegangen werden, dass diese darum bemüht, sind ihre eigene Glaubwürdigkeit möglichst plausibel unter Beweis zu stellen und somit kontraproduktive Äußerungen entsprechend zu vermeiden (vgl. GREUEL et al., 1998). UNDEUTSCH (1967) macht hier allerdings auf Ausnahmefälle aufmerksam; dies wären allenfalls sehr intelligente Personen mit hoher Lügenmotivation.

Das Merkmal ist dann erfüllt, wenn ein Zeuge im Rahmen der Exploration zwar Einwände gegen die Richtigkeit seiner eigenen Aussage vorbringt, seine Darstellung als unplausibel, unwahrscheinlich oder merkwürdig kommentiert, dabei aber trotzdem selbst von der Richtigkeit der Angaben überzeugt ist (vgl. KÖHNKEN, 2001).

GREUEL et al. (1998) unterscheidet weiter mehrere Aspekte solcher Einwände. Zum einen können dies solche Bekundungen sein, wonach ein Zeuge von sich aus die Meinung äußert, dass seine Schilderung unwahrscheinlich klinge, aber trotzdem so verlaufen sei. Zum anderen sind hier solche Äußerungen gemeint, wonach der Zeuge die Möglichkeit von fehlerhaften Wahrnehmungen, Verwechslungen oder Missverständnissen seinerseits einräumt.

Auch hier kann man eine Negativabgrenzung im Sinne eines Nicht-Erfüllt-Seins des Merkmals anführen, wie z.B. pauschale Begründungen von Zeugen, sie hätten aus einer Angst heraus, es würde ihnen niemand glauben, keine Abgaben gemacht. Hierbei handelt es sich vielmehr um die nachträgliche Begründung eines langen Schweigens, das jedoch in einer Motivationsanalyse näher ergründet werden müsste (vgl. GREUEL et al., 1998).

‘Selbstbelastende Äußerungen’ im Sinne eines Realkennzeichens sind dann in einer Aussage gegeben, wenn ein Zeuge sich selber oder sein Verhalten im Kontext des inkriminierten Geschehens in einer ungünstigen Weise darstellt, die dem eigenen Prestige abträglich ist und insofern Unvorteilhaftes impliziert und gelten nach GREUEL et al. (1998) somit als Hinweis auf den Erlebnisbezug einer Aussage bzw. lassen zumindest ein Bemühen um Sachlichkeit und Objektivität der Aussage erkennen. Hierunter fallen insbesondere Eingeständnisse von eigenen Fehlern, Angaben über die eigene Initiative, Zustimmung oder Kooperation bei den fraglichen Geschehnissen, Ermutigung des Beschuldigten oder positive Reaktionen auf dessen Verhalten. Man geht dagegen davon aus, dass ein intentional falsch aussagender Zeuge eher bestrebt ist, sich selbst im besten und den Beschuldigten im schlechtmöglichsten Licht darzustellen, als umgekehrt. Dabei muss jedoch immer im Einzelfall geprüft und bewertet werden, ob auch für beide Seiten (aus der Perspektive des Beschuldigten wie auch des Zeugen) die vermeintlich selbstbelastenden Angaben prestigemindernde Bedeutung haben (vgl. Littmann & Szewczyk, 1983 in GREUEL et al., 1998).

Besonders bei kindlichen oder jungen Zeugen beiderlei Geschlechts gelten auch solche Schilderungen als belegkräftiges Merkmal für den Erlebnishintergrund einer Aussage, die gewisse Schema-Diskrepanzen bezüglich der jeweiligen Geschlechtsrollenstereotypen aufweisen (vgl. GREUEL et al., 1998).

Analog zu den `Selbstbelastungen` ist in dem Merkmal **`Entlastung des Beschuldigten`** ein Hinweis auf eine wohlwollend-neutralen Einstellung des Zeugen gegenüber dem Beschuldigten sowie auf eine mit Belastungsmotiven nicht vereinbare, unvoreingenommene Aussagehaltung zu sehen, die somit auf den Erlebnisgehalt der Aussage schließen lässt (vgl. STELLER & KÖHNKEN, 1989 in: GREUEL et al., 1998). Dieses Merkmal ist erfüllt, wenn der Zeuge den Beschuldigten explizit entlastet oder aber auf die Möglichkeit der Belastungssaggrivation (implizite Entschuldigung der behaupteten inkriminierten Handlungen z.B. aufgrund von Alkoholeinfluss oder anderen Gründen) verzichtet. Nach GREUEL et al. (1998) hat dieses Merkmal allerdings nur dann hohe Belegkraft, wenn man davon ausgeht, dass ein erlebnisbezogen aussagender Zeuge generell eine neutrale oder wohlwollend-neutrale Haltung gegenüber dem Beschuldigten einnimmt, wohingegen ein intentional falsch aussagender Zeuge grundsätzlich einen übertriebenen Belastungseifer erkennen lässt.

Deliktspezifische Inhalte

`Deliktspezifische Aussageelemente` werden in dieser gesonderten Kategorie erfasst. Delikttypische Details sind ein sehr wichtiger Indikator in Bezug auf die Glaubhaftigkeit einer Aussage. Sie liegen dann vor, wenn ein Zeuge mit seinen Angaben ein *aufgrund der Alltagserfahrung nicht erwartbares Verhaltensmuster* schildert, das aufgrund fundierter kriminologischer, kriminalpsychologischer und/oder viktimologischer Untersuchungen in Bezug auf einen bestimmten Deliktbereich als spezifisch beurteilt werden kann, jedoch dem Zeugen selbst als solche nicht bekannt sind (vgl. KÖHNKEN, 2001). Es kann sich dabei um spezifisches Täterverhalten handeln, aber auch um opfertypische Reaktionsweisen. Beide Aspekte liegen für gewöhnlich außerhalb des alltäglichen Erfahrungsbereiches psychologischer oder kriminologischer Laien und können in der Regel deshalb von diesen in ihrer Bedeutung nicht reflektiert werden; ihr Vorhandensein in einer Aussage hat daher durch die gewonnene psychologische Stimmigkeit und deliktspezifische Folgerichtigkeit hohe Signifikanz im Hinblick auf die Glaubwürdigkeit der Angaben, die die allgemeine Aussagehomogenität übertrifft (vgl. GREUEL et al., 1998). GREUEL et al. (1998) verweist darauf, dass sich der psychologischen Forschungslage nach bestimmte Konstellationen von individuellen Verhaltensmustern oder interpersonellen Beziehungsentwicklungen (im wechselseitigen Zusammenspiel) empirisch nachweisen lassen, die von Tätern wie auch Opfern unter den spezifischen Bedin-

gungen des relevanten Deliktgeschehens auch im überindividuellen Vergleich mit hoher Auftretenswahrscheinlichkeit zu erwarten sind (vgl. GREUEL et al., 1998, S. 98).

Die Bewertung der Deliktspezifität einer Aussage ist der Literatur nach allerdings in hohem Maße vom *fachspezifischen, differenzierten Kenntnisstand des Beurteilers* abhängig, da sich das Merkmal 'Deliktspezifität' ja nicht nur auf den Bereich der Sexualdelikte bezieht, sondern auf den gesamten strafrechtlich relevanten Bereich anwendbar ist. Trotz der großen Differenzierungsmöglichkeiten, allein bezogen auf das Alter und den Entwicklungsstand von Kindern, ist dieses Merkmal gerade besonders geeignet für die Beurteilung kindlicher Zeugaussagen, speziell über Missbrauchsdelikte, da diese in der Regel nur sehr beschränkte Kenntnisse und Erfahrungen in diesem Bereich haben. Besonders in den Fällen, wo Zeugen komplexe deliktspezifische Verhaltensmuster beschreiben ohne darüber entsprechende deliktspezifische Kenntnisse zu haben oder wenn von sozial abweichenden Stereotypen berichtet wird, liegt das Merkmal in qualifizierter Ausprägung vor. Dabei geht es allerdings nicht um die Bestätigung des Vorhandenseins einzelner deliktspezifischer Angaben im Sinne von Einzeldetails, sondern vielmehr um eine oft implizite Skizzierung eines komplexen und übergeordneten Verhaltensmusters, in dem sich verschiedene deliktspezifische Details wechselseitig aufeinander beziehen und sich bedingen. Die Annahme eines bestimmten Delikttyps rein aus der (klinischen) Erfahrung steht dem Sachverständigen oder anderen begutachtenden oder beurteilenden Personen nicht zu und ist auch nicht ausreichend (vgl. GREUEL et al., 1998).

Im Rahmen einer historisch-inhaltlichen Betrachtung und des Bedeutungswandels des Glaubhaftigkeitsbegriffs soll im Folgenden lediglich auf (die letzten und aktuellen) zwei der fünf diesbezüglich eingeteilten „Phasen“ und deren Charakteristika verwiesen werden, die in unmittelbarem Zusammenhang mit der vorliegenden Arbeit stehen (vgl. GREUEL, 2001):

In einer so genannten *vierten Phase der Aussagepsychologie* in den 1980er Jahren (STELLER, 1988) galt es vor allem die Evaluation und Validierung der entwickelten Qualitätsmerkmale in entsprechenden empirischen Studien und Experimenten zu forcieren (experimentelle Validierungsphase). Zentrale Konstrukte waren hier die Aussagequalität und die Aussagezuverlässigkeit. Da aufgrund vieler durchgeführter Untersuchungen die 'UNDEUTSCH-HYPOTHESE' allgemein als bestätigt angesehen wird und das Instrument der kriterienorientierten Aussageanalyse als geeignet und zuverlässig gilt, kann konstatiert werden, dass eine diesbezügliche Konstrukt- und Methodenentwicklung vorläufig abgeschlossen ist. Trotzdem muss angemerkt werden, dass einige Personengruppen, wie etwa Lernbehinderte bei Validierungsstudien nicht oder kaum berücksichtigt wurden, also keine Studien dazu vorliegen. So schreiben STELLER und VOLBERT (1997): „... Es erscheint nicht mehr nötig, Studien zur allgemeinen Gültigkeitsprüfung der Undeutsch-Hypothese anzulegen, vielmehr sollte

speziellen Fragestellungen der Vorrang gegeben werden“ (STELLER & VOLBERT, 1997, S. 20f.). Dies soll im Rahmen der vorliegenden Dissertation unter anderem erfolgen.

Mit den 1990er Jahren hat nun eine *fünfte Phase der Aussagepsychologie* begonnen, deren Focus vor allem auf theoretischer und struktureller Integration der vielen vorliegenden Einzelbefunde liegt, da aus der veränderten Gerichtspraxis heraus verschiedene Problemstellungen an die wissenschaftliche Psychologie herangetragen werden. Eine solche integrative Wende innerhalb der Psychologie der Zeugenaussage mit einer wesentlich stärkeren Differenzierung sowohl auf methodisch-diagnostischer als auch auf konzeptioneller Ebene scheint dringend erforderlich (vgl. GREUEL, 2001).

1.2.2.3 Integrierende Glaubhaftigkeitsbeurteilung

Trotzdem STELLER & KÖHNKEN mehrfach darauf hingewiesen haben, dass die merkmalsorientierte Aussageanalyse (Criteria-Based Content Analysis, CBCA) nur einen Aspekt der psychologischen Glaubhaftigkeitsbegutachtung (Statement Validity Analysis, SVA) darstellt und eben nicht wie von vielen amerikanischen Forschern fälschlich angenommen als standardisierter „Wahrheits-Test“ und im Sinne einer kontextfreien CBCA-Forschung angewendet werden kann und darf. Vielmehr dient sie dazu, die Frage zu klären, ob und inwieweit auf dem Hintergrund von wahrnehmungs-, sozial-, gedächtnis- und entwicklungspsychologisch relevanten Voraussetzungen einer Person und ihrer Umwelt eine entsprechende Aussage phänomenale Besonderheiten erlebnisfundierter Geschehnisse aufweist, die außerhalb einer so genannten subjektiven Wachwirklichkeit nicht konstruierbar sind. Dabei sind für die Gesamtbewertung gerade und besonders dieser intraindividuelle Bezug, bestimmtes eventuell vorhandenes domänenspezifisches Wissen hinsichtlich des relevanten Sachverhalts sowie differenzierte kognitive und narrative Fähigkeiten von Bedeutung. Die vorherige Prüfung individueller intellektueller Leistungsbesonderheiten ist somit unbedingt geboten (vgl. GREUEL, 2001).

Bezieht man die gerade ausgeführten Aspekte sowie die Vielzahl der sozialen Einflussfaktoren vor und vor allem nach einem Geschehen oder Erlebnis (z.B. in Form von Falschinformationseffekten, Gesprächen mit anderen Menschen oder Zeugen, der Integration eigener oder fremder, evtl. ähnlicher Ereignisse in die Schilderung) in die Betrachtung mit ein, so ist nach GREUEL (2001) demnach eine Aussage nicht mehr nur als Leistungsprodukt des Gedächtnis zu sehen, sondern auch als soziales Phänomen.

Weiterhin dürfen die Glaubhaftigkeitsmerkmale und deren Anwendung als Merkmalssystem nicht als reines „Checklistenverfahren“ missverstanden werden, bei dem allein die Anzahl des Vorliegens bestimmter notwendiger Merkmale quasi im Sinne eines normativen „cut-off-

points“ für den Erlebnisbezug einer Aussage spricht bzw. ausreicht (vgl. VOLBERT, 2000; GREUEL, 2001). Eine schematische Anwendung der Realkennzeichen ist von daher abzulehnen, dass aus dem Vorliegen einer bestimmten Anzahl von Merkmalen im Sinne eines Schwellenwertes auf die Glaubhaftigkeit einer Aussage geschlossen wird (vgl. BUNDESGERICHTSHOF, 1999).

STELLER & VOLBERT (1999) betonen zur Beurteilung einer Aussage anhand der beschriebenen Realkennzeichen zudem Folgendes: Zwar können die festgestellten Kriterien durchaus oft als Hinweis für den Erlebnisbezug einer Schilderung gelten, jedoch ist bei Fehlen dieser Merkmale ein Umkehrschluss im Sinne einer Lüge nicht zu vertreten. Das Fehlen bestimmter Realkennzeichen kann somit auch durch andere Faktoren (wie z.B. Hemmungen, Angst, Gedächtnismängel) bedingt sein. Dementsprechend kann die inhaltsanalytische Untersuchung von Zeugenaussagen zwar als Methode zur Feststellung und Manifestation des Erlebnisgehalts von Aussagen dienen, jedoch stellt sie keine Methode zur „Lügendetektion“ dar (vgl. STELLER & VOLBERT (1999).

Die Inhaltsanalyse umfasst allein die Einschätzung des einen Aspekts der Qualität einer Aussage, für ein abschließendes Urteil über den wahrscheinlichen Erlebnisbezug einer Aussage sind weitere Determinanten zu berücksichtigen. So determiniert der „unauflösbare interaktive Bezug von Aussagequalität (inkl. Konstanz und Aussageweise), personaler Kompetenz und Aussagegeschichte (Bedingungen der Erstaussage und der weiteren Aussageentwicklung) das praktische Vorgehen der Datenerhebung und das diagnostische Schlussfolgern bei der Glaubhaftigkeitsbeurteilung von Zeugenaussagen“ (STELLER & VOLBERT, 1999, S.12). So kann beispielsweise trotz gering ausgeprägter Aussagetüchtigkeit und Aussagequalität aufgrund fehlender Validitätsminderungen ein Urteil abgegeben werden, dass die Aussage als hoch wahrscheinliche und erlebnisfundiert qualifiziert. Erst die integrative Bewertung von Aussagequalität sowie externen und internen Rahmenbedingungen der Aussage kann zu einem abschließenden Urteil über die Glaubhaftigkeit der Aussage führen und muss immer vor dem Hintergrund einzelfallspezifischer Besonderheiten unter der Formulierung einzelfallspezifischer Alternativhypothesen erfolgen. Neben den genannten, hierfür relevanten Bereichen der Psychologie und Voraussetzungen der Person, sind auf Seiten der Interviewer und Beurteiler zusätzlich weit reichende und differenzierte Kenntnisse der Kriminalpsychologie, Viktimologie und Psychodiagnostik sowie reliable und intensive Rater-Trainings notwendig, was die Angemessenheit und Aussagekraft einiger US-amerikanischer Studien zumindest anzweifeln lässt, welche oftmals die genannten Kriterien nicht erfüllten (vgl. GREUEL, 2001).

1.2.2.4 Empirische Befundlage zur Validierung der kriterienorientierten Aussageanalyse (CBCA)

Letztendlich geht es in solchen Untersuchungen darum, die Operationalisierung der 'UNDEUTSCH-HYPOTHESE' auf ihre Angemessenheit bzw. Validität hin zu überprüfen, wobei diesbezüglich vor allem der empirisch-experimentelle Zugang gewählt wird, indem die Trennschärfe einzelner Glaubhaftigkeits- oder Qualitätsmerkmale überprüft wird – wie auch in dieser Untersuchung.

Aufgrund einer inzwischen vorliegenden Reihe von Validierungsstudien kann insgesamt auf der einen Seite zwar festgestellt werden, dass sowohl in Feldstudien, die sich z.B. auf Gutachtenmaterial aus abgeschlossenen Gerichtsfällen beziehen [vgl. ESPLIN, BOYCHUK & RASKIN (1988); RASKIN & ESPLIN (1991); BOYCHUK (1991); ANSON, GOLDING & GULLY (1993); LAMERS-WINKELMAN & BUFFING (1996); HOROWITZ, LAMB, ESPLIN, BOYCHUK, KRISPIN & REITER-LAVERY (1997); LAMB, STERNBERG, ESPLIN, HERSHKOWITZ, ORBACH & HOVAV (1997); ORBACH & LAMB (1999) und CRAIG, SCHEIBE, KIRCHER, RASKIN & DODD (1999) bezüglich der Aussagen von Kindern und KRAHÉ & KUNDROTAS (1992) und PARKER & BROWN (2000) bezüglich der Aussagen von Erwachsenen] als auch in Simulationsstudien auf der Basis experimentell induzierter Aussagen [vgl. YUILLE (1988); JOFFE & YUILLE (1992); STELLER, WELLERSHAUS & WOLF (1992); HOMMERS (1997); TYE, AMATO, HONTS, DEVITT & PETERS (1999) und AKEHURST, KÖHNKEN & HÖFER (2001)] nachgewiesen werden konnte, dass sich erlebnisfundierte Aussagen qualitativ von erfundenen Aussagen ohne Erlebnishintergrund unterscheiden. Auf der anderen Seite muss konstatiert werden, dass sich sämtliche oben aufgeführten Untersuchungen auf Aussagen von normal begabten Kindern, Jugendlichen oder Erwachsenen beziehen, sich jedoch nach Kenntnis des Verfassers der vorliegenden Arbeit bis dato keine einzige Studie zur Validierung der CBCA bei lernbehinderten Jugendlichen oder Menschen mit sonstigen intellektuellen Einschränkungen vorliegt bzw. finden ließ.

Über alle genannten – methodisch doch zum Teil recht unterschiedlichen - Studien hinweg zeigt sich, dass insbesondere allgemeine und spezielle Qualitätsmerkmale bzw. Glaubhaftigkeitsmerkmale zwischen erlebnisbasierten und erfundenen Aussagen diskriminieren. Dabei handelt sich vor allem um folgende Merkmale: logische Konsistenz, quantitativer Detailreichtum, unstrukturierte Darstellung, kontextuelle Einbettung, Schilderung nonverbaler Aktionen, Schilderung von Komplikationen im Handlungsverlauf, Wiedergabe von Gesprächen, Schilderung nebensächlicher Details und Schilderung eigener psychischer und/oder physischer Vorgänge. Hinsichtlich der übrigen Einzelmerkmale fanden sich uneinheitliche, zum Teil sogar hypothesen- und erwartungskonträre Ergebnisse.

In den bisherigen Untersuchungen konnten dagegen im Hinblick auf die Kategorie der motivationsbezogenen Kriterien (wie z.B. Zugeben von Erinnerungslücken, Selbstbelastungen, ...) nur geringe bzw. keine Effekte oder Trennschärfen gefunden und festgestellt werden. Dabei stellt sich allerdings die Frage, ob und inwieweit überhaupt die motivationale Situation eines falsch aussagenden Zeugen vor Gericht experimentell induzierbar ist. Aus psychologischer Sicht dürfte sich das Befolgen einer Instruktion im Rahmen einer experimentellen Befragungssituation, möglichst gut und überzeugend zu lügen, zumindest in einigen Studien oder Gruppen von Versuchspersonen, jedenfalls erheblich von der Motivation eines intentional falsch aussagenden Zeugen unterscheiden (vgl. GREUEL, 2001). Die Übertragbarkeit der Trennschärfe der motivationsbezogenen Merkmale bleibt demnach auf jeden Fall zu diskutieren.

Da einige Aspekte für die vorliegende Arbeit von Bedeutung sind, sei an dieser Stelle zum einen auf eine Übersichtsarbeit von RAPP (1999) über verschiedene Simulationsstudien mit Kindern hinsichtlich deren Erinnerungsleistung vor allem in Abhängigkeit von unterschiedlichen Interviewtechniken verwiesen. Insgesamt ist es den Ergebnissen dieser Studien zufolge wahrscheinlich, dass bei mehrfacher (gleichartiger) Interviewbefragung, Informationen, die über mehrere Interviews wiederholt werden, recht zuverlässig sind. Dagegen sollten Informationen, die erst nach einem längeren Zeitraum zum ersten Mal berichtet werden, kritisch und mit Vorsicht betrachtet werden. Zusammengefasst scheint nach den vorliegenden Untersuchungen ein erstes Interview die Erinnerungsleistung bei einem zweiten Interview prinzipiell erhöhen zu können. Dabei spielt jedoch je nach Alter der Kinder die Art des verwendeten Interviews eine Rolle, wobei hier das Kognitive Interview durch seine Strukturvorgabe und Vorgabe von mehr externen Erinnerungsstrategien sich als hilfreich erwies. Bezüglich des juristischen Bereichs sollte beachtet werden, dass die Menge falscher Informationen über mehrere Interviews hinweg recht klein und stabil zu bleiben scheint. Tendenziell sind dabei Informationen, die zu beiden Interviews erinnert werden, korrekt, während Informationen, die beim zweiten Interview erstmals berichtet werden, wahrscheinlich falsch sind (vgl. Flin et. al, 1992; Gee & Pipe, 1995; Fivush & Schwarzmüller, 1995; Salomon & Pipe, 1997; zit. nach RAPP, 1999).

Zum anderen liegen neben der Studie von AKEHURST, KÖHNKEN & HÖFER (2001), deren mögliche Anwendbarkeit auf Aussagen von Erwachsenen ebenfalls bestätigt wurde, eine Reihe von Untersuchungen der kriterienorientierten Aussageanalyse bei Erwachsenen vor, auf die hier nur verwiesen werden kann KÖHNKEN & WEGENER (1982); KÖHNKEN & SCHIMOSSEK (1991); WELLERSHAUS & STELLER (1992); LANDRY & BRIGHAM (1992);

HÖFER, KÖHNKEN, HAHNEWINKEL & BRUHN, 1993; ZAPARNIUK, YUILLE & TAYLOR (1995); WOLF & STELLER (1997); DAHLE & SPORER (1997); RUBY & BRIGHAM (1998).

Exemplarisch soll hier auf kurz auf die Untersuchung von KÖHNKEN, SCHIMOSSEK, ASCHERMANN & HÖFER (1995) hingewiesen werden:

Die Hauptuntersuchungsfrage bestand hier darin festzustellen, ob das Kognitive Interview nach Geiselman & Fisher (1989) einen negativen Einfluss auf die kriterienorientierte Aussageanalyse zur Unterscheidung von wahren und erfundenen Aussagen von Erwachsenen habe. Dazu gaben 59 Versuchspersonen entweder einen wahren oder einen erfundenen Bericht über eine Blutspendeszene ab. Die Versuchspersonen der Wahrheitsgruppe sahen dazu einen 12-minütigen Film, der bezüglich der Schemastruktur eines solchen Ereignisses auf wissenschaftlicher Grundlage produziert wurde. Die Versuchspersonen der Phantasiegruppe erhielten entsprechend eine Liste mit verbalen Hinweisen über den Ablauf der Szene im Film. Anschließend wurden die Teilnehmer entweder mit einem kognitiven oder mit einem strukturierten Interview befragt, deren Transkripte dann zum einen anhand der Anzahl von berichteten Details (korrekt, nicht korrekt, konfabuliert) zum anderen hinsichtlich des Vorkommens von Realkennzeichen untersucht wurden.

Im Ergebnis fanden sich signifikant mehr korrekte und konfabulierte Details in den Aussagen, die auf einem kognitiven Interview basierten als in denen eines strukturierten Interviews. Die inhaltlichen Glaubhaftigkeitsmerkmale differenzierten ebenfalls zuverlässig zwischen den Versuchsgruppen. Es fand sich keine Wechselwirkung zwischen dem Wahrheitsstatus und der Interviewform, was erkennen lässt, dass das Kognitive Interview die Differenzierungsfähigkeit (zwischen wahren und erfundenen Aussagen) der kriterienorientierten Aussageanalyse nicht beeinträchtigt (vgl. KÖHNKEN, SCHIMOSSEK, ASCHERMANN & HÖFER, 1995).

Vergleicht man nun die vorangegangenen Darstellungen der verschiedenen Feld- und Simulationsstudien, so kann insgesamt festgestellt werden, dass zum einen die ökologische Validität der experimentellen Untersuchungen zu den Realkennzeichen deutlich zugenommen hat, zum anderen liegt der Schluss nahe, dass sich die systematische Auswertung der Praxis und die empirisch-experimentelle Forschung ergänzen (sollten).

Bei Betrachtung der Befunde im Einzelnen zeigt sich, dass auch in erlebnisbasierten Aussagen kaum jemals alle Qualitätsmerkmale zu finden sind, dabei jeweils auch verschiedene Merkmale statistisch signifikante Differenzen aufweisen und dass zudem die in Diskriminanzanalysen ermittelten relativen Gewichte der Realkennzeichen zum Teil recht unterschiedlich ausfallen. Da diese inhaltlichen Merkmale naturgemäß nicht unabhängig von der Art der berichteten Geschehnisse bzw. des Sachverhalts sind, erscheint der manchmal aufkommende Vorschlag, einzelne in einigen Studien nicht signifikante Merkmale aus dem Katalog der Realkennzeichen zu entfernen ebenso unsinnig wie die Auffassung von Greuel et al.

(1998), den motivationsbezogenen Merkmalen grundsätzlich eine geringere Belegkraft zuzuschreiben als denen, die sich auf die individuellen kognitiven Fähigkeiten eines Zeugen beziehen. Dazu fehle nach KÖHNKEN et al. (1995) eine entsprechende empirische Legitimation. Eine pauschalisierte Ablehnung bzw. Überschätzung bestimmter Merkmale sei deshalb unzureichend und zu vermeiden, da gerade beispielsweise motivationsbezogene Inhalte in einigen Untersuchungen sich als besonders trennscharf erwiesen haben (vgl. KÖHNKEN, 1999A).

Als mögliche Alternative oder Ergänzung zur CBCA schlugen STRÖMWALL et al. (2004) den Realitätsüberwachungsansatz ('reality monitoring - approach', RM) vor. Beide Instrumente wurden von den Autoren hinsichtlich des Ausmaßes des Einflusses der Anzahl dessen analysiert, wie oft Kinder ein Ereignis erlebt oder sich vorgestellt haben. Dabei nahm die Hälfte der Kinder (im Alter zwischen 10 und 13 Jahren) tatsächlich an einer ärztlichen Untersuchung teil (ein- bzw. viermal), während die andere Hälfte sich dies nur vorstellte (ebenfalls ein- oder viermal). In der Befragung nach einer Woche zeigte sich der RM-Ansatz als gut geeignet sowohl hinsichtlich der Authentizität der Aussagen (erhöhter Anteil der Kriterien für reale Ereignisse) als auch bezüglich der Frage, ob das Ereignis wiederholt stattgefunden hat bzw. sich vorgestellt wurde (erhöhter Anteil der Kriterien für wiederholte Handlungen). Dagegen konnte die CBCA nicht erfolgreich zwischen tatsächlichen und vorgestellten Ereignissen differenzieren (vgl. STRÖMWALL et al., 2004). Auf den Vergleich der beiden Analyseinstrumente, deren Vorteile und Nachteile, der Frage der Anwendbarkeit bei speziellen Personengruppen und Implikationen für weitere Forschungsarbeiten wird im Diskussionsteil noch näher eingegangen.

Insgesamt kann trotz der mehrfach erwähnten methodischen Unterschiede der einzelnen Studien die Gültigkeit der Undeutsch-Hypothese (sowohl bei Kindern als auch bei Erwachsenen) in ihren wesentlichen Punkten festgestellt und die CBCA als geeignetes Instrument zur Glaubhaftigkeitsbegutachtung angesehen werden. Wichtige Voraussetzung ist dabei allerdings eine gute Ausbildung und Kenntnis der Rater sowie entsprechend hohe Raten der Inter- und Intra-Rater-Reliabilität. Dies war in einigen Studien nicht unbedingt der Fall (Krahé & Kundrotas, 1992; Ruby & Brigham, 1997). Jedoch bleibt die letztendliche konkrete Gewichtung der einzelnen Merkmale bei der Gesamtbeurteilung von Aussagen ungeklärt (vgl. GREUEL et al., 1998).

Inwieweit sich die merkmalsorientierte Aussageanalyse als hilfreiches oder geeignetes Verfahren zur Einschätzung der Glaubhaftigkeit von Aussagen von Menschen (Jugendlichen) mit intellektuellen Einschränkungen erweist, bleibt zu diskutieren.

1.2.3 Juristische Aspekte

1.2.3.1 Das BGH-Urteil: Methodische und formale Standards der Glaubhaftigkeitsbegutachtung

In einem Urteil des Bundesgerichtshofes vom 30. Juli 1999 (1 StR 618/98 – LG Ansbach) wurden die derzeit verbindlichen Standards und Mindestanforderungen an aussagepsychologische Begutachtungen festgelegt, die in der fachpsychologischen Literatur bereits schon lange unter dem Aspekt der Qualitätssicherung forensisch-psychologischer Begutachtung diskutiert werden. Grundlage dafür waren die diesbezüglich von den Professores Dr. Steller und Dr. Fiedler eingeholten wissenschaftlichen Gutachten. Hintergrund ist die erfolgreiche Revision des Angeklagten in der oben genannten Strafsache wegen sexuellen Missbrauchs eines Kindes (vgl. BALLOFF, 2000).

Methodische Standards für die Planung und Durchführung einer konkreten Begutachtung im Einzelfall

Das methodische Grundprinzip der Begutachtung besteht darin, einen zu überprüfenden Sachverhalt (hier: die Glaubhaftigkeit der spezifischen Aussage) so lange zu negieren, bis diese Negation mit den gesammelten Fakten nicht mehr vereinbar ist. Somit nimmt der Sachverständige bei der Begutachtung zunächst die Unwahrheit der Aussage an (sog. Nullhypothese), welche dann unter Bildung weiterer Alternativhypothesen geprüft wird. Je nach Ergebnis und Abgleichung mit den übrigen erhobenen Fakten wird die Nullhypothese dann entweder verworfen oder beibehalten. Die Bildung von Hypothesen ist daher von ausschlaggebender Bedeutung für Inhalt und (methodischen) Ablauf einer Glaubhaftigkeitsbegutachtung (vgl. BUNDESGERICHTSHOF, 1999).

Formale Standards der schriftlichen Gutachtenabfassung

Nachdem ein psychologisches Gutachten eine komplexe wissenschaftliche Leistung darstellt, muss es sich auch entsprechend an solchen Kriterien messen lassen, die an wissenschaftliche Arbeiten angelegt werden, wobei es vor allem um die Kriterien der `Transparenz` und der `Nachvollziehbarkeit` geht. Es existieren zwar keine normativen Vorgaben über die formale Struktur von psychologischen Gutachten, jedoch gibt es durchaus fachwissenschaftlich anerkannte Standards des inhaltlichen Gutachtenaufbaus, die im Folgenden kurz aufgelistet und stichpunktartig erläutert werden sollen (vgl. GREUEL, 2000). Demnach muss ein Gutachten generell (mindestens) folgende Gliederungspunkte enthalten:

- (1) Angaben zum Begutachtungsablauf
- (2) Wiedergabe der Anknüpfungstatsachen
- (3) Deskriptive Wiedergabe der erzielten Untersuchungsergebnisse
- (4) Psychologischer Untersuchungsbefund
- (5) Beantwortung der Untersuchungsfragestellung
- (6) Zusammenfassung

Neben den beiden gerade dargestellten Schwerpunktbereichen ist an dieser Stelle auch ein dritter zu nennen, der des Konzepts der *'Glaubhaftigkeit von Aussagen'*, welcher aber schon in der vorangehenden Darstellung erläutert wurde.

Nach STELLER & VOLBERT (1999) lassen sich demnach die allgemeinen Standards (Gütemaßstäbe) der forensisch-psychologischen Diagnostik unter folgenden vier Gesichtspunkten zusammenfassend betrachten:

Zum einen ergeben sich rechtliche und ethische Standards, zum anderen geht es um theoretische Aspekte des diagnostischen Entscheidungsprozesses, um die inhaltlichen Grundlagen für die Bearbeitung spezieller Gutachtenfragen sowie um die Richtlinien der konkreten Gutachtenerstellung (Gutachtentechnik).

Demnach gelten zusammenfassend folgende wesentlichen Standards (vgl. STELLER & VOLBERT, 1999):

- Bezugnahme auf die spezifische Aussage (aussagepsychologische Indikatoren)
- Erkennbare Spezifizierung der globalen gerichtlichen Fragestellung für den Einzelfall, d.h. Formulierung von relevanten Fragestellungen und Hypothesen
- Datensammlung auf der Basis der ausgewählten Fragestellungen (hypothesengeleitete Diagnostik)
- Erkennbare Überprüfung relevanter Alternativhypothesen, wobei der Abwägungsprozess des Gutachters und dessen diagnostisches Schlussfolgern deutlich werden muss
- Verwendung einer wissenschaftlich begründeten Methodik (Analyse der Aussagequalität mit Schwerpunkt auf CBCA, Konstanzanalyse und der Analyse der Aussageweise)

1.2.3.2 Prozessrechtliche Stellung des Zeugen und des Sachverständigen

Zeugen und deren Aussagen sind ein unverzichtbares, aber auch umstrittenes Beweismittel, da sie und deren Beweiswert aufgrund der unkalkulierbaren Einflüsse aus ihrer Persönlichkeit und verschiedener situativer Faktoren als nur bedingt zuverlässig gelten. Dennoch kommt der Würdigung von Zeugenaussagen und ihren Beweiswert im Strafprozess besondere Bedeutung zu, da dies nach der Rechtsprechung ureigenste Aufgabe des Gerichts ist. Dies gilt besonders bei Gewalt- und Sexualdelikten (vgl. FREUDENBERG, 2001; BENDER & NACK, 1995). In der Strafprozessordnung (StPO) wird eine besondere Stellung des Zeugen festgesetzt. Im eigentlichen juristischen Sinn ist der Zeuge demnach ein persönliches Beweismittel, eine Beweisperson, die in einem nicht gegen sie selbst gerichteten Strafverfahren Auskunft über ihre Wahrnehmung von Tatsachen gibt, die im Zusammenhang mit dem Verfahren stehen (vgl. MEYER-GOßNER, vor § 48, Randnr. 1). Nach der StPO geht man davon aus, dass auch solche Menschen als Zeugen fungieren können, die körperlich oder geistig eingeschränkt sind, sofern sie nur zu Wahrnehmungen von Objekten, Personen oder Vorgängen und deren Wiedergabe vor Gericht in der Lage sind. Ist eine akustisch zu vernehmende Zeugenaussage nicht möglich oder zu erwarten, kann auf mimische oder gestische Artikulation zurückgegriffen werden. Eine allgemeine Zeugnisunfähigkeit gibt es also nicht (vgl. MEYER-GOßNER, vor § 48, Randnr. 12/13). BUSSE & VOLBERT (1997) weisen darauf hin, dass Kinder als sehr verletzbare Zeugengruppe angesehen werden, da sie im Vergleich zu Erwachsenen aufgrund ihres Entwicklungsstandes (Wissen, Ausdrucksmöglichkeiten, soziale Kompetenz) über weniger Bewältigungsstrategien für Belastungssituationen verfügen (vgl. BUSSE & VOLBERT, 1997). Dementsprechend ist analog ebenso bei Menschen mit Behinderungen und/oder intellektuellen Beeinträchtigungen davon auszugehen.

Der *Sachverständige* ist neben dem Zeugen in allen Verfahrensordnungen eines der persönlichen Beweismittel. Um objektiv messbare Kriterien zur Prüfung des Wahrheitsgehalts von Aussagen kindlicher und jugendlicher Zeugenaussagen zu erlangen, werden in Strafprozessen zunehmend forensische Sachverständige vor allem aus der Psychologie im Kontext forensischer Wahrheitsfindung und Urteilsbildung zu Rate gezogen (vgl. FREUDENBERG, 2001). Der Gutachter unterliegt dabei der Strafprozessordnung und wird vom Richter in seiner Tätigkeit geleitet (gemäß § 78 StPO). Grundsätzlich gelten für den forensischen, mit einer Aussagebegutachtung beauftragten Sachverständigen, wie für jeden anderen Sachverständigen auch bestimmte maßgebliche Rechtsvorschriften (im Strafverfahren die §§ 72-85 StPO). Diese(r) ist demnach verpflichtet, sein Gutachten persönlich, unparteiisch, objektiv, sachgerecht, auf dem aktuellen Stand der Erkenntnisse der jeweiligen Fachdisziplin basierend, nachvollziehbar, kontrollierbar und rechtzeitig zu erstatten. Außerdem muss er

richterliche Fragen – sein Fachgebiet betreffend – nach bestem Wissen und Gewissen beantworten und diese Antworten gegebenenfalls beenden. Darüber hinaus hat der Sachverständige auch das Recht, über den Ermittlungsstand des Verfahrens informiert zu werden, Aufklärung über begutachtungsrelevante Tatsachen zu erhalten sowie an Vernehmungen teilzunehmen. Nach GREUEL et al. (1998) besteht die erste Aufgabe des Sachverständigen darin, dem Auftraggeber eine umfassende und nachvollziehbare Antwort auf die von ihm formulierte gutachterliche Fragestellung zu geben. In Fällen von Glaubhaftigkeitsbegutachtungen bezieht sich diese Antwort immer auf die Frage des wahrscheinlichen Erlebnisgehalts und die Zuverlässigkeit einer Zeugenaussage zu einer inkriminierten Straftat. Auf der Ebene der forensisch-psychologischen Diagnosestellung ist nun konkret anzugeben und im Gutachten darzustellen, wie die Glaubhaftigkeit der Aussagen in Bezug auf jeden einzelnen Schilderungssachverhalt einzuschätzen ist (vgl. GREUEL et al., 1998, S. 287 f.).

1.3 Intelligenzminderung und Lernbehinderung

1.3.1 Terminologie

Trotzdem eine exakte Beschreibung der Störungsbilder und die Einschätzung der kognitiven und anderen Leistungskapazitäten (auch in Abhängigkeit des angelegten Kriteriums) der an der vorliegenden Untersuchung teilnehmenden Versuchspersonen nicht ganz einfach ist und die Grenzen zwischen verschiedenen Begrifflichkeiten fließend sind, soll im Folgenden dennoch eine Klassifikation der hier relevanten Begriffe sowie Abgrenzungen voneinander und Analogien und Differenzen zum angloamerikanischen Sprachraum vorgenommen werden.

1.3.1.1 Definitionen nach WHO (World Health Organization) und ICD-10 (International Classification of Diseases)

Die WHO beschreibt und definiert *Intelligenzminderung* nach ICD, dem Klassifizierungssystem psychischer und psychiatrischer Störungen, als „eine sich in der Entwicklung manifestierende, stehen gebliebene oder unvollständige Entwicklung der geistigen Fähigkeiten (verstanden), mit besonderer Beeinträchtigung von Fertigkeiten, welche zum Intelligenzniveau beitragen, wie zum Beispiel Kognition, Sprache, motorische und soziale Fähigkeiten“ (DILLING, H., MOMBOUR, W., SCHMIDT, M. H., 1993, S. 254). Weiterhin liegt stets eine Anpassungsstörung des Verhaltens vor, die jedoch bei Personen mit leichter Intelligenzminderung in geschützter Umgebung mit Unterstützungsmöglichkeiten nicht unbedingt auffällt. Außerdem besteht für diese Personengruppe ein größeres Risiko ausgenutzt sowie körperlich und sexuell missbraucht zu werden (vgl. DILLING, H., MOMBOUR, W., SCHMIDT, M. H., 1993). Die Einteilung des Schweregrads der Intelligenzminderung erfolgt in erster Linie nach Werten des IQ, wobei diese im Hinblick auf die internationale Vergleichbarkeit nicht allzu starr gewertet werden sollten. Je nach Schweregrad sind Menschen mit Intelligenzminderung in der selbstständigen Lebensgestaltung, in der Unabhängigkeit ihrer Versorgung, im Erlernen schulischer und beruflicher Fertigkeiten und auch in ihrer sozialen und emotionalen Entwicklung beeinträchtigt.

Nach MÖLLER et al. (2001) ist *Intelligenz* ein menschliches Merkmal, das sich in der Bevölkerung annähernd normalverteilt. Etwa 5 % der Bevölkerung ist von einer Minderung der Intelligenz betroffen, wobei 3 - 4 % der Bevölkerung leichte Formen aufweisen und etwa 1 % von einer schweren Form der Intelligenzminderung betroffen ist (vgl. MÖLLER, H. J., LAUX, G. & DEISTER, A. (2001).

Im Folgenden soll die Einteilung der Intelligenzminderung dargestellt werden, wobei nicht nur die Einteilung nach dem Intelligenzniveau und nach möglichen Defiziten im Vordergrund stehen soll, sondern ebenfalls Kompetenzen aufgezeigt werden sollen.

Tabelle 1-4: Einteilung der Intelligenzminderung

IQ	Nomenklatur	IQ	Nomenklatur
69 bis 50	leichte Intelligenzminderung	49 bis 35	mittelgradige Intelligenzminderung
34 bis 20	schwere Intelligenzminderung	unter 20	schwerste Intelligenzminderung

(vgl. MÖLLER, H. J., LAUX, G. & DEISTER, A., 2001, S.406)

Im Bereich der *leichten Intelligenzminderung* ist häufig eine Verzögerung im Spracherwerb feststellbar. Normale Konversation sowie klinische Interviews sind jedoch meist problemlos möglich. Schwierigkeiten sind vor allem auch im schulischen Bereich (vor allem Lesen und Schreiben) vorhanden, die allerdings bei ausreichender Förderung soweit trainiert werden können, dass eine weitgehende Eingliederung ins berufliche und gesellschaftliche Leben erfolgen kann. Organische Ursachen sind bei einer Minderheit der Betroffenen festzustellen. Begleiterkrankungen wie andere Entwicklungsverzögerungen oder Störungen des Sozialverhaltens kommen in unterschiedlicher Anzahl vor.

Bei einer *mittelgradigen Intelligenzminderung* zeigen sich sehr unterschiedliche Leistungsprofile. So kann die Fähigkeit bestehen, einer einfachen Unterhaltung zu folgen und sich an dieser zu beteiligen oder es können Sprachkompetenzen zur Mitteilung von Grundbedürfnissen verfügbar sein. Auch die Fähigkeit zum Erlernen lebenspraktischer Fertigkeiten kann verzögert bis nicht vorhanden sein. Im schulischen Kontext können grundlegende Fähigkeiten im Lesen, Schreiben und Rechnen erlernt werden, so dass später unter Beaufsichtigung durchaus eine Verrichtung einfacher praktischer Tätigkeiten möglich ist. Das schulische Vorankommen ist allerdings begrenzt. Bei der Mehrzahl der Betroffenen finden sich Anzeichen für eine soziale Entwicklung, beispielsweise in der Fähigkeit, Kontakt aufzunehmen, mit anderen zu kommunizieren und einfache soziale Aktivitäten zu erbringen. Ebenfalls können bei der Mehrzahl der Personen eine organische Ursache sowie bei einem Teil des Personenkreises andere tief greifende Entwicklungsstörungen beobachtet werden. Das allerdings niedrige Sprachniveau erschwert oft eine Diagnosestellung.

Menschen, die von einer *schweren Intelligenzminderung* betroffen sind, leiden an einer deutlich ausgeprägten motorischen Schwäche oder anderen Ausfällen, welche auf das Bestehen einer klinisch bedeutsamen Schädigung oder Fehlentwicklung des Zentralnervensystems hinweisen.

Von einer *schwersten Intelligenzminderung* betroffene Personen sind so gut wie unfähig, Aufforderungen oder Anweisungen zu verstehen oder sich danach zu richten und somit nur zu sehr rudimentären Formen nonverbaler Kommunikation fähig (vgl. DILLING, H., MOMBOUR, W., SCHMIDT, M. H., 1993).

Nach SCHOPF (1998) gibt die WHO mit der Einführung des ICDH (International Classification of Impairments, Disabilities and Handicaps) nur einen sehr allgemeinen Rahmen für die Klassifikation von Lernbehinderung vor, der mit spezifischen Inhalten des jeweiligen Bereiches gefüllt werden muss (vgl. SCHOPF, 1998).

Angaben zu Häufigkeit und Prävalenz schwanken in Deutschland und auch international erheblich zwischen etwa 0,4 % und 3 % der Gesamtbevölkerung. In Studien wurden im Altersbereich zwischen 7 und 16 Jahren Prävalenzraten um ca. 0,7 bis 0,8 % bei Kindern und Jugendlichen gefunden. Sicher ist, dass geistige Behinderung bei Jungen etwa 1,5-mal häufiger als bei Mädchen auftritt (vgl. SENN, 1993).

Nach ZELFEL (2001) ist in Deutschland davon auszugehen, dass es in der Bevölkerung etwa 800.000 bis 1 Million Menschen gibt, die mit dem unscharfen Begriff „lernbehindert“ bezeichnet werden können (vgl. ZELFEL, 2001).

1.3.1.2 Konzepte, Modelle und begriffliche Abgrenzungen:

Lernbehinderung - Lernschwäche - Lernbeeinträchtigung - geistige Behinderung

Um sich dieser nicht ganz eindeutigen, aber wichtigen begrifflichen Thematik in einem ersten Schritt anzunähern, scheint es angebracht, zunächst einmal den Begriff der *Behinderung* zu umschreiben.

Nach dem deutschen Sozialgesetzbuch (SGB) gilt eine Person als behindert, wenn sie in Folge einer körperlichen Schädigung des Organismus, einer Schwäche der geistigen Kräfte oder einer seelischen Störung nicht nur vorübergehend daran gehindert ist, Funktionen und Aktivitäten so auszuüben, wie sie innerhalb einer Bandbreite als normal betrachtet werden, und somit bei der Ausfüllung der für die Person im übrigen (nach Alter, Geschlecht, sozialem Kontext...) als normal angesehene Rolle in der Gesellschaft benachteiligt ist (vgl. § 2 (1) SGB IX, 2004, S. 6).

Dabei treten wiederum evtl. uneindeutige Begrifflichkeiten auf, die einer weiteren Klärung bedürfen (vgl. RAT für KRIMINALITÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN, 2003).

Behinderung:

Beeinträchtigung im Alltag aufgrund einer Schädigung

Beeinträchtigung:

..., wenn wegen der Schädigung die Anforderungen der natürlichen und sozialen Umwelt (...) nicht oder nur eingeschränkt erfüllt werden können (Aktivitätseinschränkung)... oder bei einer Maßnahme, Struktur oder Verhaltensweise anderer, die geeignet ist, wegen einer Schädigung die Teilhabe am Leben der Gemeinschaft oder bei der selbstbestimmten Lebensgestaltung zu erschweren, einzuschränken oder zu verhindern (Teilhabe einschränkung)

Schädigung:

Die nicht nur vorübergehende (mindestens sechs Monate) Einschränkung einer körperlichen Funktion, geistigen Fähigkeit oder seelischen Gesundheit.

Nach den durch die Bundesanstalt für Arbeit erlassenen Durchführungsanweisungen ist gemäß § 19 SGB III geregelt, dass zum Personenkreis der Behinderten auch *lernbehinderte Menschen* gehören, unabhängig von der besuchten Schule. Dabei gelten als Lernbehinderte „junge Menschen, die in ihrem Lernen *umfänglich und lang andauernd beeinträchtigt* sind und die deutlich von der Altersnorm abweichende Leistungs- und Verhaltensformen aufweisen, wodurch ihre berufliche Integration wesentlich und auf Dauer erschwert wird“ (SCHOPF, 1998, S. 403). Die beiden Begriffe „umfänglich“ und „lang andauernd“ sind dabei zentral, wobei sich ersterer auf die Folgen der Schädigung oder Störung bezieht und vor allem der individuelle und soziale Kontext bewertet wird, letzterer wiederum in Zusammenhang mit dem zeitlichen Kontext und dem Verlauf steht.

Ursprünglich entstanden die Begriffe „lernbehindert“ und „Lernbehinderung“ um 1960 im Zusammenhang mit der Umbenennung der Hilfsschulen in Schule für Lernbehinderte. Eine klare inhaltliche Fassung und Abgrenzung von anderen Begriffen wie „Lernstörung“ oder „Schulleistungsschwäche“ scheint nicht einfach, da die damit in Verbindung stehenden Phänomene nicht immer direkt offensichtlich sind (vgl. WEIß). Dies hat verschiedene Gründe; zum einen, da dieser eine Vielzahl von unterschiedlichen Beobachtungen und Wertungen umschreibt (vgl. ZELFEL, 2001), zum anderen da in diesem Zusammenhang immer individuelle Stärken, Ressourcen, Anlagen und Neigungen zu berücksichtigen sind ebenso wie das Lebensalter, sekundäre Behinderungen sowie die Wechselwirkungen dieser Einflussfaktoren (vgl. SCHOPF, 1998).

Eine wissenschaftlich präzise Definition scheint auch aufgrund kaum vorhandener exakter wissenschaftlicher Forschungsergebnisse schwierig, dennoch werden oft umschreibende Begriffe genutzt, um ein bekanntes Phänomen zu skizzieren:

- Schüler(innen) mit (einer) Lernbeeinträchtigung(en) bzw. mit Beeinträchtigungen des Lernens
- Kinder und Jugendliche mit Beeinträchtigungen in bestimmten Entwicklungsbereichen sowie des Lern- und Leistungsverhaltens
- Schüler(innen) mit sonderpädagogischem Förderbedarf

Die Begriffe Lernbehinderung und Benachteiligung sind unscharfe Begriffe und wenig klar definiert, da kaum Kriterien für die Diagnose, die Schweregrade und die Abgrenzung zu anderen Behinderungen und Störungen (Verhaltensstörung, Verhaltensauffälligkeit, Lernschwäche, Lernbeeinträchtigung, Lernstörung) bestehen – obwohl die Lernbehinderung quantitativ betrachtet an erster Stelle der Formen und Arten von Behinderungen steht.

Nach einer Art 'Ausschlussprinzip' nennt WEIß einige Aspekte, die im Zusammenhang mit der problematische Beschreibung von Lernbehinderung, zumindest auszuschließen sind:

- Lernbehinderung ist keine „umfassende Behinderung einer allgemeinen Lernfähigkeit“, da es eine solche nicht gibt und jeder Mensch in verschiedenen Lernbereichen individuell unterschiedlich ausgeprägte Fähigkeiten entwickelt.
- Lernbehinderung ist keine individuelle Eigenschaft, die als Ursache der Lernschwierigkeiten (...) oder gar als Wesensmerkmal bestimmter Kinder angesehen werden könnte.
- Lernbehinderung ist schließlich nicht einfach mit einer Intelligenzschwäche gleichzusetzen; denn lernbehinderte Schülerinnen und Schüler weisen in ihren Intelligenzleistungen eine hohe Streuung von unter 60 bis über 100 IQ-Punkten auf.

Wie in anderen Fachbereichen auch hat sich bei der Diagnose 'Lernbehinderung' folgende Einteilung bewährt, wobei Störungen aus allen drei Bereichen zugrunde liegen können (vgl. SCHOPF, 1998):

(1) *spezifische neurologische Störungen*

z. B.: Anfallsleiden, motorische Verlangsamung, Wahrnehmungsstörungen, Störungen der Reagibilität, etc.

(2) *Teilleistungsstörungen/-schwächen*

(sie betreffen die „höheren“ kulturellen Leistungen sowie die komplexeren Leistungen des Gehirns in den Bereichen Motorik, Sprache, Denken, Schreiben, Lesen, etc.)

z. B.: schwankende Aufmerksamkeit meist mit Konzentrationsstörungen, verringerte Unterscheidungsfähigkeit, Probleme beim Umsetzen von verbalen Anweisungen, etc.

(3) *Basisstörungen*

(unspezifische Störungen, die nicht unbedingt einem System zuzuordnen sind, sondern mehrere oder alle Funktionen tangieren)

z. B.: verzögerte Wahrnehmung; erhöhte Ablenkbarkeit; starke Schwankungen und allgemeine Labilität der Leistung; verringerte Lernleistung; Verlängerung von Handlungen, Entscheidungen, Versuch und Irrtum; labile Motivation; etc.

Lernbehinderungen sind somit vielgestaltig und können auch durch vielerlei Einflüsse zustande kommen, aufrechterhalten und verstärkt werden. Somit kann und soll die komplexe Leistung des Lernens kaum bzw. nicht nur auf einzelne Bestimmungsfaktoren reduziert werden. Intelligenzmängel und defizitäre kognitive Verarbeitungsprozesse wirken sich sicherlich auf Lernprozesse aus, haben aber dennoch eine eigene Entstehungsgeschichte und können nicht als Hauptursache gesehen werden.

Reduziert man jedoch die Einordnung des Begriffs rein auf die *kognitive Ebene*, so spricht man im IQ-Bereich zwischen 70 und 90 von einer Lernbehinderung (vgl. MÖLLER, H. J., LAUX, G. & DEISTER, A., 2001).

Konkret sind im Zusammenhang mit der Störung der kognitiven Entwicklung Lernbehinderter vor allem folgenden Bereiche relevant bzw. problematisch (vgl. SCHOPF, 1998; FOLLOWAY et al., 1997):

(1) *Schwäche der Abstraktionsfähigkeit* (die höheren kognitiven Funktionen sind nicht Leistungen einzelner Strukturen, sondern des Zusammenspiels). Deshalb finden sich unter anderem Schwierigkeiten...

- bei abstrakten Begriffen
- Größen und Benennungen zu unterscheiden
- Vergleiche durchzuführen und Größen zu schätzen
- übergeordnete Kategorien zu bilden und Typen, Klassen und Symbole zu verstehen
- symbolische Verknüpfungen durchzuführen
- etwas zu generalisieren
- bei Transferleistungen
- etwas unter einem anderen Blickwinkel zu sehen

(2) *die konkretistische Realität des Lernbehinderten*

Dinge sind weniger Objekte, sondern mehr konkrete Subjekte, Worte weniger Symbole als konkrete Namen. Dies zeigt sich auch im Verhalten. Das Handeln erfolgt daher nicht aus einem übergeordneten Handlungskonzept (Selbstbild, Fremdbild, übergeordnete Ziele, Werte), sondern aus sehr unmittelbaren, emotionalen Bedürfnissen. Das Aufschieben von Befriedigungen zum Erreichen höherwertiger Ziele ist ihm oftmals nicht möglich und mitursächlich für die Lern- und Motivationsstörungen

(3) *starre Verhaltensmuster*

Manche höheren kognitiven Fähigkeiten werden gar nicht oder nur lückenhaft entwickelt, was sich in Konsequenz in rigiden Lösungsstrategien und Handlungsschemata manifestiert, die nun als starre Schemen auf alle ähnlichen Situationen angewandt werden.

(4) *Metakognition*

Kennzeichnend ist hier die Schwäche von 'höheren' Sinn- und Bedeutungskonzepten für die Steuerung des eigenen Verhaltens. Diese Störungen behindern daher den Lernprozess entscheidend und schwächen die Aneignung der erforderlichen Kompetenzen. Es zeigt sich eine gravierende Diskrepanz zwischen Lernanforderungen und den durch die Behinderung bedingten Lernstil lernbehinderter Menschen.

Informationsverarbeitung findet auf konkretem Niveau statt im Sinne einfacher (kindlicher) Lernformen. Lernbehinderte neigen dazu, alte Muster und Programme starr zu wiederholen, da keine Strategien vorhanden sind, sich Neues anzueignen und einzuüben.

Carroll (1981; zit. nach BORKOWSKI & BÜCHEL (1983) nennt 10 wichtige und fundamentale kognitive Prozesse, von denen mehrere bei Retardierten mangelhaft ausgebildet bzw. gestört sind:

- Monitoring - begleitende Sorgfalt - perzeptuelle Integration - Enkodierung - Vergleichen - darstellende Fähigkeiten und Behalten - Übertragung – Antwortausführung -

In einer Übersicht stellen ZIGLER & BALLA (1982) verschiedene Erklärungsansätze für die insgesamt langsamere und qualitativ andere kognitive Entwicklung von lernbehinderten Personen dar:

- (1) Nach *Defektttheorien bzw. der Strukturdifferenzhypothese* werden Verarbeitungsdefizite bei Lernbehinderten durch neurophysiologische Defekte bestimmt.
- (2) Die *Defizittheorie bzw. Entwicklungsverzögerungshypothese* nach ZIGLER sieht bei Intelligenzschwachen eine langsamere und früher zum Stillstand kommende Entwicklung bei sonst normaler Entwicklungsfolge.
- (3) In neueren Untersuchungen würden nun vermehrt *Informationsverarbeitungsmodelle* im Vordergrund stehen, wobei der Fokus hierbei auf kognitiven Funktionen wie Aufmerksamkeit, Gedächtnis, Wissen(erwerb) sowie Umgang mit Begriffen und Problemlösefertigkeiten liegt.

Nach MÄHLER & HASSELHORN (1990) sind im Rahmen der so genannten 'developmental-difference-Kontroverse' zwei sich gegenüberstehende Auffassungen hinsichtlich kognitiver Minderleistungen zu unterscheiden.

Nach der *Entwicklungsverzögerungshypothese* ist Lernbehinderung als eine Verzögerung der normalen kognitiven Entwicklung zu verstehen. Unterschiede zu normal begabten bestünden demnach vor allem in der Entwicklungsrate und dem maximal erreichbaren Entwicklungsniveau.

Demgegenüber postuliert die *Strukturdifferenzhypothese* darüber hinaus vor allem eine andersartige kognitive Struktur. Nach den Befunden neuerer Arbeiten geht man immer mehr zu Modellen der Informationsverarbeitung bzw. des Gedächtnisses über. Umstritten ist dabei die so genannte 'similar-structure-Hypothese', wonach die kognitive Struktur und Leistung von Behinderten vergleichbar sein soll mit der normal entwickelter Kinder gleichen Intelligenzalters (vgl. MÄHLER & HASSELHORN, 1990).

Neben dem Unterschied im IQ-Bereich unterscheidet sich die Lernbehinderung von einer geistigen Behinderung oder Intelligenzminderung ($IQ < 70$) durch das geringere Missverhältnis zwischen den Möglichkeiten eines Individuums und den Erwartungen seiner Umwelt, weshalb auch der Grad der Abhängigkeit von psychosozialen und pädagogischen Hilfeleistungen bei lernbehinderten Menschen im Allgemeinen geringer ist als bei geistig behinderten Menschen (vgl. WEIß).

In den Bereichen des Entwicklungsverlaufs und des Entwicklungspotentials kann nach SCHOPF (1998) festgestellt werden, dass oftmals kein altersgemäßes Sozialverhalten entwickelt wird und diese Defizite mit neurotischen Verhaltensweisen kompensiert werden. Durch soziale Ausgrenzung und Arbeitslosigkeit besteht für viele Jugendliche kaum die Möglichkeit, Selbstbewusstsein und Selbstwert durch eigene Leistung zu entwickeln. Die Ich-Entwicklung stagniert auf kindlichem Niveau. BRYAN & BRYAN (1990) gehen davon aus, dass zumindest einige Kinder oder Jugendliche mit Lernbehinderungen einem höheren Risiko ausgesetzt sind, Schwierigkeiten hinsichtlich ihres peer-status und anderer sozialer Rollen zu entwickeln und soziale Situationen zu verstehen, da ihnen in einem gewissen Ausmaß bestimmte sozial-kognitive Fähigkeiten fehlen. Dies ist vor allem in Situationen der Fall, wo die sozialen Hinweisreize mehrdeutig oder nur implizit sind oder der Jugendliche eine bestimmende Rolle in einer sozialen Interaktion übernehmen muss. Die genaue Etiologie und der ursächliche Zusammenhang sind noch nicht genau geklärt, jedoch geht man davon aus, dass die betreffenden Jugendlichen weniger Gelegenheiten haben, soziale Erfahrungen zu sammeln und demzufolge weniger soziales Wissen entwickeln bzw. Erfahrungen, wie (andere) Menschen ihre Gefühle, Gedanken und Absichten ausdrücken. Dementsprechend sind diese Personen auch weniger dazu in der Lage eigene Vorstellungen zu entwickeln oder fremde zu erkennen und zu interpretieren, unter anderem weil diese Gefühle und Stimmungen oft auch nonverbal kommuniziert werden. Solche Defizite führen dann oft

dazu, dass die Jugendlichen ungeschickt oder missverständlich mit ihren 'peers' interagieren, was wiederum negative Reaktionen provozieren kann (vgl. BRYAN & BRYAN, 1990).

Allgemein kann postuliert werden, dass die zur Verfügung stehende psychische wie physische Energie die Lernleistung bestimmt und damit die Steigung der Lern- oder Entwicklungskurve. Der behinderte Mensch ist mehrfach belastet durch:

- Störungen der Entwicklung in früheren Lebensabschnitten resultieren in Reduktion Rückständigkeit im gesamten Niveau
- Durch permanente Kraftaufwendungen zur Kompensation seiner Behinderung (psychisch, physisch, sozial und emotional)
- Durch eine behinderungsbedingte Reduzierung seiner Lernleistung (Hirnschädigung, Intelligenzminderung, etc.)

Die reduzierte Lernleistung erzeugt eine Schere mit ungünstiger Dynamik: der Abstand zur normalen Entwicklung wird immer größer.

Somit ist Lernbehinderung also nicht als statisches Wesensmerkmal eines Kindes oder Jugendliche zu sehen, sondern sie entwickelt sich – unter bestimmten Belastungen – in einem dynamischen Prozess. Zu diesen Belastungen gehören im Wesentlichen drei Faktorengruppen, die an dieser Stelle vorerst nur kurz genannt werden sollen:

- (1) *entwicklungs- und lernerschwerende biologische Faktoren* - v. a. (unbestimmte, leichtere, diffuse) Funktionsstörungen des Zentralnervensystems (Teilleistungsstörungen)
- (2) *entwicklungs- und lernerschwerende Umwelteinflüsse* (Armut, soziale Benachteiligung, niedriger Bildungs- und Erwerbsstatus der Eltern, wenig ermutigende Erziehungspraxis, unzureichende Befriedigung grundlegender kindlicher Bedürfnisse, ...)
- (3) *ungünstige schulische Lehr-/Lernbedingungen*

Demnach entsteht Lernbehinderung zu einem ganz überwiegenden Teil in Wechselwirkung mit sozialer Benachteiligung und Randständigkeit (vgl. WEIß)

Nach Angaben der „Learning Disabilities Association of Amerika“ (LDA) kann die Definition von Lernbehinderung wie folgt lauten (sinngemäße Zusammenfassung nach ZELFEL, 2001):

Lernbehinderungen sind ein bezeichnender Begriff, der sich auf eine heterogene Gruppe von Störungen bezieht, die durch identifizierbare oder unklare Dysfunktionen des zentralen Nervensystems verursacht sind. Solche Störungen können sich bemerkbar machen durch Entwicklungsverzögerungen und/oder Schwierigkeiten in den folgenden Bereichen: Aufmerksamkeit, Gedächtnis, schlussfolgerndes Denken, Koordination, Kommunikation, Lesen, Schreiben, Rechtschreibung, Rechnen, soziale Kompetenz und emotionale Reife. (ZELFEL, 2001, S. 7f.)

Weiterhin schreibt ZELFEL (2001), dass Lernbehinderungen intrinsisch sind und das Lernen und Verhalten des Individuums beeinflussen können (vgl. ZELFEL, 2001).

Verschiedene Gründe oder Einflüsse wie das sog. 'Normalisierungskonzept' sowie die Idee eines gesellschaftlichen Umdenken im Sinne einer Abkehr der Vorstellung behinderter Menschen als funktionseingeschränkte Objekte hin zu einer Integration und Ermöglichung der weitgehenden Teilhabe am gesellschaftlichen Leben (als gesellschaftliche Aufgabe) haben zu einer Überarbeitung des ICIDH geführt. In dieser neuen, seit 2000 vorliegenden Fassung *ICIDH-2 (International Classification of Functioning and Disability)* beinhaltet der Behinderungsbegriff drei Dimensionen:

(1) *Körperfunktionen und –strukturen*

Störungsbegriff: Schädigung (Funktionsstörung, Strukturschaden)

(2) *Aktivitäten* (konkret durchführbar)

Störungsbegriff: Beeinträchtigung der Aktivitäten (Aktivitäts- oder Leistungsstörung)

(3) *Partizipation*

Störungsbegriff: Beeinträchtigung der Partizipation

Somit integriert das ICIDH-2 zwei Modelle (Konstrukte) zur Erklärung von Funktionsfähigkeit und Behinderung:

- das medizinische Modell (Behinderung als persönliches Problem)
- das soziale Modell (Behinderung als gesellschaftlich verursachtes Problem und nicht als Eigenschaft der Person, sondern als komplexes Geflecht von Bedingungen)

Nach ZELFEL (2001) unterscheiden sich Menschen mit Lernbehinderungen in vielen Bereichen nicht von Behinderten; im Hinblick jedoch auf die Teilhabe am gesellschaftlichen Leben wie etwa Zugang zu Bildung und Berufsausbildung unterscheiden sie sich deutlich von anderen nicht behinderten Menschen (vgl. ZELFEL, 2001; OPP, 1992).

Im Rahmen eines *multifaktoriellen Modells* macht SCHOPF (1998) deutlich, dass die Fähigkeiten eines jeden Menschen von einer Vielzahl von Einflüssen und Rahmenbedingungen abhängen, wie beispielsweise

- den genetischen bedingten Anlagen,
- der in ihm angelegten Persönlichkeitsstruktur,
- dem sozialen Umfeld und dem Klima, in das die Entwicklung eingebettet ist,
- den erforderlichen Förderungen und Unterstützungen in den Entwicklungsphasen,
- den individuellen sich herausbildenden Verhaltensmustern (auch Fehlhaltungen).

Diese multifaktorielle Sicht der Störungen macht somit auch deutlich, dass Lernbehinderung keine einzelne Störung sein kann, sondern immer ein ganzes Gefüge von Einflussfaktoren betroffen ist, welche sich in individuellen und typischen Mustern überlagern, worin oft das Spezifische dieser Behinderung zu sehen ist.

Der Schweregrad einer Behinderung hängt dabei ab von:

- dem Schweregrad einer Störung im Vergleich zur Leistung Gleichaltriger einer Normalpopulation,
- der Vielzahl von leicht bis mittelgradigen Störungen,
- der negativen Wechselwirkung, welche leichte bis mittelgradige Störungen in ihrer Wirkung verstärkt,
- zusätzlichen Begleit- oder Sekundärbehinderungen sowie
- Auswirkungen auf die Lebensentwicklung

Somit kann die Lösung des Problems 'Lernbehinderung' bzw. deren konkreteres Verständnis nur durch eine *multidisziplinäre, dynamische Diagnostik* gelingen, die auch die vielfältigen Wechselwirkungen zwischen den einzelnen Leistungsbereichen berücksichtigt (vgl. SCHOPF, 1998).

An dieser Stelle soll auf einen von ZELFEL (2001) gemachten *Definitionsvorschlag von Lernbehinderung* hingewiesen bzw. dieser relativ ausführlich wiedergegeben werden, der nach Meinung des Verfassers dieser Arbeit die bisher dargestellten Aspekte gut zusammenfasst und beschreibt.

Lernbehinderung ist ein allgemeiner Begriff für eine heterogene Gruppe von Störungen, vielfach verbunden mit bestimmaren und unbestimmaren Dysfunktionen des zentralen Nervensystems. Solche Störungen können sich darstellen in Entwicklungsverzögerungen und/oder Beeinträchtigungen in jedem der folgenden Bereiche:

Wahrnehmungsleistung, Konzentration, Gedächtnis, schlussfolgerndes Denken, psychomotorische Koordination, Kommunikation, Lesen, Schreiben, Rechtschreibung, Rechnen, soziale Kompetenz und emotionale Reife.

Lernbehinderungen können Lernen und Verhalten bei jedem Individuum beeinträchtigen, eingeschlossen solcher mit durchschnittlicher und unterdurchschnittlicher Intelligenz. Lernbehinderungen sind nicht primär verursacht durch visuelle, auditive und motorische Behinderungen, durch retardierte geistige Entwicklung, emotionale Störungen oder umweltbezogene Nachteile, obwohl sie mit jedem dieser Faktoren auftreten können.

Lernbehinderungen können entstehen durch genetische Variationen, biochemische Faktoren, Ereignisse im vor-, zwischen- und nachgeburtlichen Bereich und/oder in Kombination mit ungünstigen Sozial- und Umweltbedingungen.

Lernbehinderungen beeinträchtigen die Aktivität des betroffenen Individuums, vornehmlich in den folgenden Bereichen:

Lernen und Wissensanwendung, im Verstehen und Produzieren von geschriebenen und gesprochenen Mitteilungen, im komplexen personellen Interagieren mit Menschen in einer kontextuell und sozial geeigneten Weise sowie in der Bewältigung von Aufgaben, von Aufgabenkomplexen und solcher Aktivitäten, die in wichtigen Lebensbereichen wie Arbeit, Bildung, Ausbildung und Freizeit erforderlich sind.

Die Teilhabe von Menschen mit Lernbehinderungen ist in folgenden Bereichen erschwert:

Informationsaustausch, soziale Beziehungen, Bildung und Ausbildung, Erwerbsarbeit und Beschäftigung, Wirtschaftsleben sowie im sozialen und staatsbürgerlichen Leben.

Menschen mit Lernbehinderungen bedürfen besonderer Förderung und der Berücksichtigung ihrer behinderungsbedingten Belange, um eine gesellschaftliche Teilhabe zu ermöglichen (vgl. ZELFEL, 2001).

In Abgrenzung zu obigen Ausführungen bezüglich des Konzeptes 'Lernbehinderung' seien der Vollständigkeit wegen noch die Begriffe der 'Lernschwäche' und der 'Lernbeeinträchtigung' kurz beschrieben:

Bei einer *Lernschwäche* liegt eine teilweise (nur einzelne Lebensbereiche betreffende), befristete Schwäche oder Hemmung vor, gemessen an den Leistungen Gleichaltriger einer Normalpopulation (leichte und zeitlich befristete Reduzierung des Entfaltungsradius sowie eine Abflachung der Lernentwicklungskurve).

Eine *Beeinträchtigung* stellt die Kategorie zwischen Schwäche und Behinderung dar, zeigt fließende Übergänge zu einer schweren Schwäche und zur leichten Behinderung und kann in zwei Formen auftreten:

- die Störung ist leichtgradig (nicht lang anhaltend), jedoch in ihrer Wirkung umfänglich
- die Störung ist mittelgradig (lang anhaltend) und nicht umfänglich, d.h. isoliert

Bei ungünstigen Bedingungen (schwächendes Milieu, keine Förderung) besteht die Gefahr einer Verstärkung der Problembereiche und einer daraus resultierenden Behinderung (vgl. SCHOPF, 1998).

Für den Begriff der *geistigen Behinderung* schlägt der deutsche Bildungsrat von 1973 folgende Definition vor:

Als geistig behindert gilt, wer infolge einer organisch-genetischen oder anderweitigen Schädigung in seiner psychischen Gesamtentwicklung und seiner Lernfähigkeit so sehr beeinträchtigt ist, dass er voraussichtlich lebenslanger sozialer und pädagogischer Hilfen bedarf. Mit der kognitiven Beeinträchtigung gehen solche der sprachlichen, sozialen, emotionalen und motorischen Entwicklungen einher. (BLEIDICK, 1977, S. 56)

Im klinischen aber auch im angewandten Bereich der Psychologie wird in erster Linie auf den Intelligenzquotienten zurückgegriffen, um verschiedene Schweregrade einer geistigen Behinderung herauszuarbeiten. Hier sei auf die schon dargestellten Ausführungen der Terminologie der Intelligenzminderung verwiesen, wonach die WHO anhand des ICD-10 Intelligenzminderungen bzw. geistige Behinderung gleichsetzt und folgende, schon beschriebene Unterteilung vorschlägt:

- | | |
|--|-----------|
| 1. Leichte Intelligenzminderung (leichte geistige Behinderung) | IQ: 50-69 |
| 2. Mittelgradige Intelligenzminderung (mittelgradige geistige Behinderung) | IQ: 35-49 |
| 3. Schwere Intelligenzminderung (schwere geistige Behinderung) | IQ: 20-34 |
| 4. Schwerste Intelligenzminderung (schwerste geistige Behinderung) | IQ < 20 |

(vgl. DILLING, H., MOMBOUR, W., SCHMIDT, M. H., 1993)

Da der Deutsche Bildungsrat die obere Grenze einer geistigen Behinderung bei einem IQ von 55 ansetzt, während das ICD-10 bei einem Intelligenzquotienten von weniger als 70 von geistiger Behinderung spricht, muss in diesem Zusammenhang kritisch angemerkt werden, dass sich somit durchaus Schwierigkeiten ergeben können, wenn man Definitionsansätze auf den kognitiven Aspekt reduziert, zumal Intelligenz nicht als ein über das Leben hinweg stabiles Maß erachtet werden kann, sondern durchaus Schwankungen unterliegt.

Insgesamt kann jedoch bei geistig behinderten Menschen dennoch davon ausgegangen werden, dass die intellektuellen Leistungsschwächen gravierender ausfallen als bei lernbehinderten Menschen und daher auch ein erhöhter Bedarf an Förderung und Betreuung notwendig ist. Zudem scheinen bei der Definition der Ursachen einer geistigen Behinderung mehr organische Schädigungen des Gehirns im Vordergrund zu stehen, wohingegen bei der Beschreibung der lernbehinderten Population auch immer psychosoziale Faktoren beachtet und in die Betrachtung miteinbezogen werden sollten.

1.3.1.3 Analogien und Differenzen zum angloamerikanischen Sprachraum

Da ein Großteil der aktuell vorliegenden Literatur zum Thema der vorliegenden Arbeit oder verwandten Themen aus dem angloamerikanischen und englischen Sprachraum stammt, soll im Folgenden kurz darauf eingegangen werden, damit unter anderem im Verlauf der weiteren Ausführungen die Begriffe bzw. der untersuchte Personenkreis besser unterschieden und eingeordnet werden können.

Die WHO (World Health Organisation) unterscheidet zunächst zwischen verschiedenen Dimensionen von Behinderung (vgl. RAT für KRIMINALITÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN, 2003):

Impairment:

medizinisch diagnostizierbare Schädigung des Organismus

Disability:

aus dem 'impairment' resultierende Einschränkung, Funktion und Aktivitäten so auszuüben, wie sie innerhalb einer Bandbreite als normal für Menschen betrachtet werden (Funktionsbeeinträchtigung).

Handicap:

Benachteiligung, die die Erfüllung der je nach Alter, Geschlecht, sozialen und kulturellen Faktoren für das Individuum sonst entsprechenden „normalen“ Rolle begrenzt oder verhindert (soziale Beeinträchtigung).

Damit stellt Behinderung nicht nur eine physische und psychische Realität dar, sondern bezieht insbesondere bzgl. des 'handicap' auch den sozialen Aspekt des gesellschaftlichen Anschlusses, der Verhinderung sozialer Kontakte und Orte mit ein.

Die gängigsten und dem im deutschen Sprachraum wohl noch am besten entsprechenden Begriffe für *Lernbehinderung* sind im Englischen: „learning disability“, „intellectual disability“ und „mental handicap“ (vgl. NEUHÄUSER, 1999).

Nach FOLLOWAY et al. (1997) wird die Definition von „learning disabilities“ des *National Joint Committee on Learning Disabilities (NJCLD)* als die brauchbarste und sinnvollste angesehen. Dabei wird folgender Definitionsversuch vorgeschlagen:

Learning disabilities is a general term that refers to a heterogenous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning or mathematical abilities. The disorders are intrinsic to the individual, presumed to be due to central nervous system dys-function, and may occur across the life span. Problems in self-regulatory behaviors, social perception and social interaction may exist with learning disabilities but do not by themselves constitute a learning disability. Although learning disabilities may occur concomitantly with other handicapping conditions (...) or with extrinsic influences (...) they are not a result of those conditions of influences"

(NJCLD, 1994, zitiert nach FOLLOWAY et al., 1997, S.299).

Eine geistige Behinderung wird in der Regel als *'mental disorder'* bezeichnet.

Der zweite, vor allem in USA gebräuchliche und näher zu beleuchtende Begriff, ist der der *'mental retardation'*, welcher in Großbritannien allerdings als diskriminierend abgelehnt wird und eher durch den Begriff *'mental handicap'* ersetzt ist. Dieser gilt wiederum in USA als herabsetzend. Bereits hier deutet sich erneut ein unterschiedliches und somit nicht eindeutiges Verständnis der Begrifflichkeiten im internationalen Gebrauch an.

Eine Definition des Begriffs *'mental retardation'* erfolgte 1992 durch die American Association on Mental Retardation und lautet:

„Mental retardation refers to substantial limitations in present functioning. It is characterized by significantly subaverage intellectual functioning, existing concurrently with related limitations in two or more of the following applicable adaptive skill areas: communication, self – care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure and work. Mental retardation manifests before age 18“.

(LUCKASSON et al., 1992, S.1).

Neben unterdurchschnittlichen intellektuellen Fähigkeiten sind Defizite im adaptiven Verhalten und eine Manifestation vor dem 18. Lebensjahr maßgeblich für diesen Definitionsansatz. Auch wenn eine Bestimmung von geistiger Behinderung über den Intelligenzquotienten unzureichend ist, so erlaubt das psychometrische Maß doch einen Vergleich der Begriffe „geistige Behinderung“ und „mental retardation“. Daher sollen kurz die Kategorien von „mental retardation“ nach GROSSMAN (1983, S.13) aufgezeigt werden, um dann eine Abgrenzung zum deutschen Begriff vornehmen zu können.

Tabelle 1-5: Kategorien von 'Mental Retardation' (MR)

IQ	Nomenklatur	IQ	Nomenklatur
ca. 70 bis 50-55	mild MR	50-55 bis 35-40	moderate MR
35-40 bis 20-25	severe MR	unter 20-25	profound MR

(aus BECKER, 1995: Sexuelle Gewalt gegen Mädchen mit geistiger Behinderung, S.42)

Nach Festsetzung der oberen Grenze der geistigen Behinderung bei einem IQ-Wert von etwa 55 durch den Deutschen Bildungsrat, entsprechen geistig behinderte Menschen in Deutschland somit im amerikanischen Sprachraum dem Personenkreis, der als 'moderate', 'severe' oder 'profound mentally retarded' eingestuft wird, also einen IQ unter 55 aufweist. Der Begriff '*mental retardation*' umfasst alle Grade intellektueller und sozialer Schwächen, wohingegen mit dem Begriff „geistige Behinderung“ nur die schwereren Formen gemeint sind (WENDELER, 1993). Des weiteren zeigt sich, dass internationale Statistiken über '*Mental Retardation*', die sich am Intelligenzniveau ausrichten, auch einen Teil des Personenkreises erfassen, der in Deutschland im Schulalter zu der Gruppe der Lernbehinderten gezählt wird und im Erwachsenenalter überhaupt nicht statistisch in Erscheinung tritt (vgl. NEUHÄUSER, 1999). Lernbehinderte Menschen in Deutschland entsprechen zudem den Menschen mit einer '*mild mental retardation*' in den USA.

Eine scharfe Grenzziehung zwischen den einzelnen Begrifflichkeiten scheint ähnlich wie im deutschen Sprachgebrauch und auch im allgemeinen Verständnis der jeweils gerade gemeinten Form von Behinderung oder Beeinträchtigung meist nur schwer zu treffen.

1.3.2 Sexueller Missbrauch und sexuelle Gewalt gegen Menschen mit Intelligenzmin- derung

Da in der öffentlichen Diskussion der letzten Jahre immer wieder auf Gewalt und Straftaten und Übergriffe sexueller Natur gegen Menschen mit Behinderung in den unterschiedlichsten Bereichen, Institutionen sowie zu verschiedenen Gelegenheiten hingewiesen wird und zudem bzw. auch demzufolge die Kriminalitätsangst und Verunsicherung von Betroffenen und Angehörigen zunimmt, soll im folgenden Kapitel überblicksartig darauf eingegangen werden (vgl. auch JOYCE, 2003).

Da sich das genannte Phänomen nicht nur auf sexuelle Handlungen bezieht, sondern auch andere Formen von Gewalt implizieren können, soll der Gewaltbegriff im Allgemeinen ebenso in die Darstellung miteinbezogen werden.

Neben der Darstellung von Studien zum Ausmaß und zur Prävalenz des Problems des sexuellen Missbrauchs von Menschen mit eingeschränkten kognitiven Fähigkeiten, sollen einige Bedingungsfaktoren dafür analysiert werden.

Zunächst sei auf die einzelnen *Begrifflichkeiten* 'Missbrauch' und 'Gewalt' bzw. deren Abgrenzung voneinander verwiesen.

Gewalt im strafrechtlichen Sinn zielt neben dem offenkundigen Aspekt der Einwirkung körperlicher Kraft insbesondere auch auf die körperliche Zwangswirkung beim Opfer (z. B. Einwirkung auf das Nervensystem und die Psyche). Somit kann von Gewalt bereits bei psychisch vermitteltem Zwang zur Überwindung eines geleisteten oder erwarteten Widerstandes gesprochen werden.

In Abgrenzung zum Gewaltbegriff liegt *Missbrauch* bereits dann vor, wenn der Täter eine Lage oder seine Stellung bzw. das Verhältnis zum Opfer hin ausnutzt, so dass ein Einverständnis des Opfers den Missbrauch nicht ausschließt, was insbesondere für den hier relevanten Personenkreis gilt, der oftmals aus verschiedenen Gründen sein Einverständnis nicht wirksam erklären kann. Aspekte des Benutzens des Opfers, das durch eine Herabwürdigung zum Objekt fremd definierter sexueller Motive wird, werden hier insbesondere angesprochen. *Sexueller Missbrauch* fällt somit unter die Definition von Gewalt, da ein Missbrauch immer mit Zwang und Machtausübung zu tun hat (vgl. RAT für KRIMINALITÄTSVERHUETUNG SCHLESWIG-HOLSTEIN ,2003).

1.3.2.1 Empirische Befunde und Untersuchungen zu Prävalenz und Inzidenz

Grundsätzlich kann über die tatsächliche Häufigkeit des Vorkommens von sexuellem Missbrauch an Menschen mit geistiger Behinderung nur spekuliert werden, da keine genauen Statistiken vorliegen und die entsprechende Forschung oftmals keine allgemeingültigen Schlüsse zulässt. So ist die untersuchte Stichprobe oft sehr klein und weist keine Kontrollgruppe auf, was keine repräsentativen Folgerungen ermöglicht. Auch das, was unter geistiger Behinderung oder Lernbehinderung verstanden wird, wie Missbrauch definiert wird sowie demographische Daten variieren von Untersuchung zu Untersuchung, was eine Generalisierung der Ergebnisse erschwert. Im Rahmen der Literaturrecherche konnte bzw. musste festgestellt werden, dass in Deutschland bislang keine repräsentative Studie zu diesem Thema veröffentlicht ist. Trotzdem können einige besondere Risikofaktoren wie intellektuelle Beeinträchtigung, Erziehung zu Anpassung und Unauffälligkeit, Hilfsbedürftigkeit und Abhängigkeit, geringes Selbstwertgefühl, emotionale Vernachlässigung, verminderte Artikulationsfähigkeit, vermeintlich geringere Glaubwürdigkeit, Abwertung, Leugnung und Reglementierung sexueller Bedürfnisse und sexuelle Unaufgeklärtheit herausgestellt werden, auf die im nachfolgenden Kapitel genauer eingegangen wird.

Das tatsächliche Ausmaß dieses Problems kann somit nur angedeutet, nicht aber empirisch belegt werden. Dennoch kommen diverse Studien überein, dass es sich um ein weit verbreitetes Problem handelt.

Insgesamt existieren im deutschsprachigen Raum zwei Studien, die einen Hinweis darauf geben können, wie weit verbreitet sexueller Missbrauch bei Menschen mit einer Intelligenzminderung ist.

So führten Noack und Schmid 1994 eine bundesweite Befragung zu dem Thema sexuelle Gewalt bei Menschen mit einer geistigen Behinderung durch. 874 Fragebögen wurden an Einrichtungen der Behindertenhilfe verschickt. Die Auswertung der 308 zurückerhaltenen Fragebögen ergab, dass 51,3 % der Mitarbeiter/-innen in den befragten Einrichtungen Fälle von sexueller Gewalt an geistig behinderten Menschen bekannt waren (vgl. NOACK & SCHMID, 1994). Dabei wurden insgesamt 574 Fälle sexuellen Missbrauchs angegeben. Opfer sexualisierter Gewalt waren demnach 53 bekannte Fälle bei weiblichen Kindern (10,7 %), 40 bekannte Fälle bei weiblichen Jugendlichen (8,7 %), 350 bekannte Fälle bei weiblichen Erwachsenen (31,5 %) und 16 bekannte Fälle bei männlichen Kindern (5,2 %), 10 bekannte Fälle bei männlichen Jugendlichen (2,9 %) und 105 bekannte Fälle bei männlichen Erwachsenen (16,6 %). Jedoch scheint im Vergleich mit internationalen Forschungsergebnissen die Dunkelziffer in deutschen Einrichtungen nach wie vor sehr hoch zu sein. Die Untersuchung bestätigte zudem, dass Mädchen und Frauen erheblich öfter Opfer sexuellen Missbrauchs werden als Jungen und Männer (vgl. NOACK & SCHMID, 1994).

Daneben ergab eine von Zemp und Pircher 1996 in Österreich durchgeführte explorative Fragebogenerhebung, die Aufschluss zum Ausmaß und Häufigkeit sexuellen Missbrauchs bei Frauen mit einer geistigen Behinderung geben sollte, ähnliche Ergebnisse. Der Befragung lag eine weit gefasste Definition sexualisierter Gewalt zugrunde. Die Autorinnen bevorzugten den Begriff der sexuellen Ausbeutung. Nach ihrer Auffassung gehört neben körperlicher Gewalt auch sexuelle Belästigung und unangenehme Berührungen zu sexueller Ausbeutung. Sie befragten 130 Frauen mittels eines Fragebogens. Die Frauen waren durchschnittlich 34 Jahre alt und lebten in der Mehrheit in Einrichtungen. 57,5 % der Befragten hatten ausschließlich eine geistige Behinderung, 23,3 % waren körperbehindert und 12,5 % waren mehrfach behindert. Dabei gaben 62,3 % der Frauen an, sexuell belästigt worden zu sein und 64 % gaben sexualisierte Gewalterfahrungen an. (vgl. RAT für KRIMINALITÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN, 2003.) In 39,4 % der Fälle war der Täter den betroffenen Frauen bekannt. In 23,1 % war der Täter unbekannt und es handelte sich Zufallsbegegnungen. In 13,3 % der Fälle von sexueller Gewalt war der Täter ein Heimbewohner. Pflege- oder Stiefväter wurden in 6,1 % der Fälle als Täter angegeben und zu 5,1 % übte das Betreuungspersonal sexuelle Gewalt aus. Die Auswertung der Fragebögen ergab, dass ca. 41 % der Frauen mehrmals in ihrem bisherigen Leben sexuelle Gewalt erfahren hatten. Frauen, die schon in Einrichtungen aufgewachsen sind, hätten signifikant mehr sexuelle Gewalt erfahren. (vgl. ZEMP & PIRCHER, 1996).

Diese explorative Untersuchung kann Hinweise auf das Ausmaß von sexueller Gewalt bei Menschen mit einer geistigen Behinderung im deutschsprachigen Raum geben, jedoch wird nichts desto trotz eine repräsentative Studie benötigt, die in ihrer Methodik den wissenschaftlichen Standards genügt. Nach SENN (1993) ließe sich das Ausmaß des Problems erfolgreich untersuchen, wenn man quer durch die verschiedenen Arten der Unterbringung dieser Menschen eine repräsentative Auswahl zusammenstellen würde und auf diesem Weg versucht, eine Stichprobe mit der größtmöglichen Bandbreite aller Grade von Behinderungen auszuwerten (vgl. SENN, 1993).

Der Anteil der leicht bis schwer intelligenzgeminderten Personen (IQ unter 80) wird auf etwa 4 % der Gesamtbevölkerung geschätzt (WALLER, 1997; KVAM, 1998). Aus einer von der Kriminalpolizei Schleswig-Holstein herausgegebenen Statistik zu Sexualstraftaten im Jahr 1995 in Schleswig-Holstein ist ersichtlich, dass von 199 mutmaßlich von sexuellem Missbrauch betroffenen Schulkindern 27 die Förderschule besuchten, dies entspricht 13,5 % der angezeigten Fälle (SCHÜLER, 1997). Zu vergleichbaren Ergebnissen kommen FEGERT et al. (2001). In der von ihnen untersuchten Stichprobe von 57 Kindern, bei denen ein sexueller Missbrauch als gesichert galt, wiesen 14,6 % eine Intelligenzminderung bzw. Lernbehinderung (IQ 83 - 74) auf. In einer groß angelegten Studie untersuchten SULLIVAN & KNUTSON (2000) zwei komplette Schülerjahrgänge (insgesamt 50278 Kinder) und stellten fest, dass die Prävalenz von Missbrauch/Misshandlung bei den Kindern ohne Behinderung bei 9 % lag und bei Kindern mit Behinderung bei 31 %.

Die genannten Befunde stützen die bereits früher von MILNE et al. (1996), BULL (1995), GUDJONSSON et al. (1993), FLIN et al. (1992), DENT (1986) und der WHO (vgl. DILLING et al., 1993) geäußerte Einschätzung, dass bei Personen mit Intelligenzminderung ein höheres Risiko als bei normal begabten Personen vorliegt, Opfer einer Straftat (Misshandlung oder Sexualdelikte) zu werden und somit als Zeugen bei Polizei oder Gericht aussagen zu müssen.

BECKER (1995) fasst verschiedene im angloamerikanischen Sprachraum schon früher und vermehrt durchgeführte Untersuchungen zu dieser besonderen Missbrauchsthematik zusammen. Hier findet man einen relativ umfangreichen Überblick über Forschungsarbeiten zu dieser Thematik. Einige dieser Untersuchungen sollen hier kurz angeführt werden.

In einem Überblick von Linda Moreau wird die Lebenssituation geistig behinderter Menschen, die zu Hause leben, aufgezeigt. Moreau kommt zu dem Ergebnis, dass 80 % dieser Menschen körperlich oder sexuell misshandelt wurden, Beschimpfungen ausgesetzt waren oder beraubt wurden (vgl. BECKER, 1995).

In einer weiteren kanadischen Studie von Joanne Doucette (1986) wurden 30 Frauen mit einer Behinderung über ihre Misshandlungs- und Missbrauchserfahrungen in der Kindheit

befragt. Kontrollgruppe waren 32 nicht behinderte Frauen. Doucette kommt zu dem Ergebnis, dass 47 % der behinderten Frauen im Vergleich zu 34 % der nicht behinderten Frauen in ihrer Kindheit sexuell missbraucht wurden. Damit haben Mädchen mit Behinderungen im Vergleich zu nicht behinderten Mädchen ein um 38 % erhöhtes Risiko, sexuelle Gewalt zu erleben, so die Autorin (vgl. BECKER, 1995).

CHAMBERLAIN et al. (1984) befragten 87 jugendliche Mädchen mit kognitiven Einschränkungen im Alter von 11-23 Jahren. 25 % der Mädchen berichteten, eine Vergewaltigung oder einen Vergewaltigungsversuch erlebt zu haben, wobei sieben junge Frauen sexuellen Missbrauch durch eine nahe Bezugsperson (Vater, Stiefvater, Pflegevater) schilderten. Die Mädchen waren zur Zeit der Übergriffe durchschnittlich 14 Jahre alt, so dass es sich um sexuellen Kindesmissbrauch handelte. CHAMBERLAIN et al. bemerken, dass ein Drittel der Mädchen mit leichten kognitiven Defiziten Opfer sexueller Übergriffe wurden, was die Vermutung nahe lege, dass diese Personengruppe besonders gefährdet sei. Möglicherweise unterschätzt diese Untersuchung jedoch das tatsächliche Ausmaß des Problems der sexuellen Gewalt, da nur Vergewaltigungsversuche und Vergewaltigungen in die Definition des sexuellen Übergriffs aufgenommen wurden. Auch stammten die Informationen aus Elternbefragungen und Akteneinsichten, was die Frage aufwirft, ob so alle Fälle von sexuellem Missbrauch erhoben werden konnten (vgl. CHAMBERLAIN, 1984).

Stromsness (1993) interviewte 27 Menschen mit einer leichten Intelligenzminderung bezüglich einer Missbrauchsthematik und ihrer sexuellen Aufklärung. Zusammen berichteten 11 Frauen von insgesamt 59 Fällen des sexuellen Missbrauchs. Im Vergleich dazu lagen 22 Fälle bei 7 Männern vor. Die geschilderten Übergriffe waren in der Mehrzahl solche, wo ein direkter sexueller Kontakt stattfand (Anal-, Vaginal, Oralverkehr). In 64 % der Fälle wurde den Frauen ein Schweigegebot vom Täter abgenommen. Die Ergebnisse von Stromsness (1993) weisen auf eine hohe Rate des sexuellen Missbrauchs bei Menschen mit einer leichten Intelligenzminderung hin (vgl. STROMSNESS, 1993).

Nach SENN (1993) seien die Forschungsergebnisse der Untersuchung von Hard, (1986, zitiert nach SENN, 1993) repräsentativ zur Einschätzung des Ausmaßes des sexuellen Missbrauchs an Menschen mit geistiger Behinderung. Hard interviewte 65 entwicklungsretardierte Personen im Alter zwischen 17 und 48 Jahren mit Hilfe eines Fragebogens. Dabei wurde sexueller Missbrauch als sexueller Kontakt wie Geschlechtsverkehr, Anal- oder Oralverkehr oder Berührungen von Brust oder Genitalien definiert, der dadurch zustande kam, dass die sexuell missbrauchten Personen unwissend oder ohne Verständnis für die sexuellen Handlungen waren oder der Täter Gewalt, Zwang oder Manipulationen ausübte. Es ergab sich, dass 83 % der Frauen und 32 % der Männer sexuellen Missbrauch erfuhren, wobei von den 38 Personen, bei denen das Alter zur Zeit des Missbrauchs bekannt war, 45

% unter 18 Jahre alt gewesen waren, wobei sich zudem herausstellte, dass 99 % der Täter aus dem sozialen Nahraum der Opfer stammten.

Untersuchungen aus den USA (BALOGH et al. (2001) und Norwegen (KVAM, 2000) kamen durch Erhebungen bei Kindern und Jugendlichen mit Entwicklungsverzögerungen bzw. Intelligenzminderung zu dem Schluss, dass die Prävalenzrate von Straftaten gegen die sexuelle Selbstbestimmung in dieser Population (unter Zugrundelegung einer eher engen Definition) bei etwa 45-50 % liege. In einer Erhebung in der Bundesrepublik (KLEIN, WAWROK & FEGERT, 1999) lag die Prävalenzrate der von Sexualstraftaten Betroffenen in der untersuchten Population von Jugendlichen mit Intelligenzminderung bei ca. 30 %. Problematisch an diesen Erhebungen erscheint allerdings die Wahl der Außenkriterien, da nicht nur als substantiiert angesehene Fälle mit einbezogen wurden, sondern auch versucht wurde, quasi das Dunkelfeld durch Fremdurteile der behandelnden Ärzte oder Einschätzungen der Betreuungspersonen mit abzuschätzen. Dennoch ergibt sich bei BALOGH et al. (2001), wenn nur die als substantiiert eingeschätzten Fälle zugrunde gelegt werden, eine Prävalenzrate von 29 %.

Die hier genannten Untersuchungen sind nur eine Auswahl. Alle weiteren Studien, die aus dem angloamerikanischen Bereich stammen, weisen weitgehende Probleme auf. Häufig werden die Eltern der Kinder und Jugendlichen befragt, dabei ist jedoch zu bedenken, dass Eltern wohl kaum zugeben werden, das ihr Kind von ihnen persönlich sexuell missbraucht wurde oder von einem nahen Angehörigen. Bei direkten Befragungen der Menschen mit einer geistigen Behinderung treten oft Kommunikationsschwierigkeiten auf. Fragebögen sind nicht adäquat konstruiert. Zum Teil wird in den Studien nicht nach Geschlecht differenziert, was vor allem im Bezug auf sexuellen Missbrauch von Nöten ist, da ungleich mehr Mädchen als Jungen betroffen sind.

So kann zusammenfassend festgestellt werden, dass im angloamerikanischen Bereich Untersuchungen zum sexuellen Missbrauch an Menschen mit einer intellektuellen Beeinträchtigungen zu finden sind. Jedoch lassen sich in fast allen Studien methodische Mängel feststellen, was eine Generalisierung fragwürdig erscheinen lässt.

Des Weiteren ergibt sich schon aufgrund des begrifflichen Unterschieds von „mental retardation“ in den USA und Kanada und „geistiger Behinderung“ oder „Lernbehinderung“ in Deutschland, die Schwierigkeit, die Forschungsergebnisse aus dem amerikanischen Raum auf Deutschland zu übertragen. Wie bereits erwähnt, kann jedoch angenommen werden, dass das Ausmaß des Problems im deutschen Sprachraum ähnlich gelagert ist.

Aus den genannten Ergebnissen ist ebenso ersichtlich, dass nicht nur geistig behinderte Kinder und Jugendliche von sexueller Gewalt betroffen sind, sondern in erhöhtem Ausmaß

auch geistig behinderte Erwachsene, was den Schluss nahe legt, diesen Personenkreis in die Überlegungen zur sexuellen Gewalt an geistig behinderten Menschen einzubeziehen und weitere Forschungsarbeiten hinsichtlich dieser speziellen Personengruppe anzuregen.

Trotzdem das Thema des sexuellen Missbrauchs bei Menschen mit einer Intelligenzminde- rung nun doch zunehmend von Interesse zu sein scheint oder zumindest des Öfteren auch von der Öffentlichkeit wahrgenommen wird, ist es doch überraschend, dass kaum empirische Datenerhebungen stattfanden. Die Arbeitsgruppe „14“ aus Schleswig-Holstein stellt fest, dass selbst Dunkelfeldforschungen nicht bekannt sind (vgl. RAT für KRIMINALITÄTS- VERHUETUNG SCHLESWIG-HOLSTEIN, 2003).

So bemerkt LUCKASSON (1992), dass wohl weit mehr Menschen aus diesem Personen- kreis betroffen sind, als angenommen wird, da die Rate der bekannt gewordenen Miss- brauchsfälle unter der durchschnittlichen Aufklärungsrate liegt, was nach LUCKASSON (1992) auf verschiedene Faktoren zurückzuführen sei, die für das Schweigen der Opfer verantwortlich sein könnten. Neben der Tatsache, dass sich Menschen mit einer Behinde- rung zumeist in Abhängigkeitsverhältnissen bewegen und sie Konsequenzen von nahesteh- enden Personen befürchten, besteht immer noch die Gefahr, dass Aussagen von Men- schen mit einer Intelligenzminderung nicht genügend ernst genommen bzw. für unglaublich befunden werden (vgl. LUCKASSON, 1992; FOLLOWAY et al., 1997).

Trotz obig bemerkter methodischer Einschränkungen kann auch als Fazit der Untersuchun- gen aus dem deutschen Sprachraum festgehalten werden, dass behinderte Menschen mit einem erhöhten Risiko leben, Opfer von sexuellem Missbrauch zu werden.

Aufgrund bestehender Kommunikationsschwierigkeiten, begrenztem Zugang zu Hilfsangebo- ten und auch mangelnder Kompetenz, die notwendigen Folgeschritte bei sexuellem Miss- brauch einzuleiten, lässt sich vermuten, dass die Dunkelziffer von sexueller Gewalt gegen Menschen mit Behinderung noch viel höher ist (vgl. auch JOYCE, 2003).

1.3.2.2 erhöhte Vulnerabilität von Menschen mit kognitiven Einschränkungen – Analyse von Bedingungs- und Risikofaktoren und Erklärungsansätze

Der Gewaltbegriff per se wurde durch die Vereinten Nationen wie folgt definiert:

„Unter Gewalt ist jede gewaltsame Handlung zu verstehen, die einen möglichen oder tatsächlichen physischen, sexuellen oder psychologischen Schaden hervorruft, einschließlich Drohungen, Vernachlässigungen, Ausbeutung, Zwang, willkürlichem Freiheitsentzug, sowohl im öffentlichen Leben als auch im Privatleben.“

(RAT für KRIMINALITÄTSVERHUETUNG SCHLESWIG-HOLSTEIN, 2003)

In erster Linie besteht in der psychologischen Fachliteratur weitgehende Übereinstimmung darin, dass Kinder, Jugendliche und auch Erwachsene mit kognitiven Defiziten einem erhöhten Risiko ausgesetzt sind, physisch misshandelt zu werden. Begründet wird das Risiko damit, dass die Belastung der Eltern, für ein behindertes Kind zu sorgen, Misshandlungen begünstige und unrealistische Erwartungen seitens der Eltern an ihre Kinder, welche nicht erfüllt werden können, zu Frustration führe (vgl. SENN, 1993). Insofern muss davon ausgegangen werden, dass Menschen mit intellektuellen Beeinträchtigungen einem mindestens genauso hohen Risiko ausgesetzt sind, wie andere Personen.

Traditionelle Erklärungsversuche eines erhöhten Missbrauchsrisikos für Personen mit Beeinträchtigungen basieren nach SENN (1993) in erster Linie auf einer 4-Stufen-Theorie, wonach:

- Beeinträchtigung in vermehrter Abhängigkeit resultiert,
- Abhängigkeit in erhöhtem Stress für den Versorger oder Verantwortlichen,
- Stress in Frustration und dem Verlust von Hemmung und
- der Verlust von Hemmung in Missbrauch.

Allerdings gilt dieses Modell als empirisch nicht bestätigt und scheint sogar in Widerspruch mit dem aktuellen, vorliegenden experimentellen Forschungsstand zu stehen (vgl. SENN, 1993).

Insgesamt betrachtet lässt sich jedoch kein direkter Zusammenhang zwischen sexuellem Missbrauch und Behinderung sehen, sondern vielmehr ein Zusammenhang zwischen sexuellen Gewalterfahrungen und verschiedenen gesellschaftlichen und sozialen Rahmenbedingungen, welchen behinderte Menschen ausgesetzt sind. LINDSAY & BEAIL (2004) verweisen in ihrem Artikel auf verschiedene Kategorien von Risikofaktoren wie z.B. statische, geschichtliche, aktuelle Variablen sowie dynamische, unmittelbare und direkte Variablen.

Einige Risikofaktoren, die den Zusammenhang zwischen sexuellem Missbrauch und kognitiven Einschränkungen oder Behinderungen erklären könnten, sollen im Folgenden dargestellt und erläutert werden (vgl. RAT für KRIMINALITÄTSVERHÜTUNG SCHLESWIG-HOLSTEIN, 2003; BECKER, 1995). Dabei können allerdings keine konkreten Aussagen darüber getroffen werden, in welchem Ausmaß die genannten Aspekte zu einem erhöhten Risiko beitragen, Opfer zu werden. Die Reihenfolge der verschiedenen Aspekte stellt somit keine Hierarchisierung da.

Mangelnde Bildung und fehlende Information

Dies meint vor allem den verwehrten Zugang zu verschiedenen Bildungsmöglichkeiten sowie eine Einschränkung durch bauliche, strukturelle oder kommunikative Barrieren. Bezüglich eventueller Rechtsansprüche und Gesundheitsfragen findet sich oft eine Vorenthaltung allgemeiner Informationen.

Soziale Isolation

So konnte beispielsweise Finkelhor (1980, zitiert nach BECKER, 1995) zeigen, dass soziale Isolation der Familie einen Risikofaktor für sexuellen Missbrauch darstellt, indem er Studenten und Studentinnen an amerikanischen Colleges und Universitäten nach sexueller Gewalt in der Kindheit befragte. Von den Frauen, die auf Farmen aufwuchsen, erlebten 44 % in ihrer Kindheit sexuelle Gewalt.

So wäre es durchaus denkbar, dass Familien mit behinderten Kindern möglicherweise sozial isolierter leben als andere Familien, da sich diese eventuell ihrer Angehörigen schämt und demzufolge vom öffentlichen Leben fernhält. So können beispielsweise stigmatisierende und vorurteilsbehaftete Verhaltensweisen von außen stehenden Personen zur sozialen Isolation beitragen (BECKER, 1995). Dadurch wären der Aufbau und die Pflege von sozialen Kontakten durch Mobilitätseinschränkungen sowie Barrieren erschwert (vgl. RAT für KRIMINALTÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN, 2003).

Insofern zeigt sich die soziale Isolation als Risikofaktor in besonderem Maße bei Familien mit behinderten Kindern.

Des Weiteren kann auch die Lebenssituation behinderter Kinder kann zu einer zusätzlichen Isolation führen. Oftmals sind diese Kinder und Jugendlichen in Ganztagsbetreuungsstätten und -schulen mit großem Einzugsgebiet untergebracht, wodurch eine Isolation von Gleichaltrigen im heimischen Wohngebiet erfolgt. Die Gefahr der Ausbildung einer Außenseiterrolle ist also gegeben. In Einzelfällen bzw. einzelnen Einrichtungen sind zwar soziale peer-Kontakte unter den Jugendlichen möglich und auch feststellbar, jedoch kann auf dem Hintergrund nicht immer davon ausgegangen werden.

Regelmäßiger Kontakt zu Freunden und Freundinnen würde untereinander soziale Kontrolle ermöglichen können, was positiv zu sehen ist dergestalt, dass Mitglieder der 'peer-group' beispielsweise durch Missbrauch hervorgerufene Veränderungen im Verhalten der betroffenen Person erkennen und dem Opfer als Ansprechpartner Hilfestellung leisten könnten. Bestehen solche regelmäßigen Kontakte nicht, mangelt es auch an sozialer Kontrolle. So zeigt sich, dass auch die Isolation von Gleichaltrigen einen Risikofaktor für sexuellen Missbrauch darstellt und die Möglichkeit, sexuelle Gewalt aufzudecken und zu beenden, weiter erschwert ist.

Persönliche und wirtschaftliche Macht- und Abhängigkeitsverhältnisse

Bei Menschen mit Intelligenzminderung bzw. Behinderung kann von einem mehrfachen Abhängigkeits- und Gewaltverhältnis ausgegangen werden. So besteht zum einen ein Machtgefälle zwischen Erwachsenen und Kindern, zum anderen ein ungleiches Geschlechterverhältnis zwischen Männern und Frauen. Bei sexualisierter Gewalt und Übergriffen gegen Menschen mit Behinderungen, welche im Übrigen vor allem im sozialen Nahraum der Opfer stattfinden, ist als dritter Faktor ein ungleiches Machtverhältnis zu nennen, dass aus der Abhängigkeit behinderter Menschen von Pflege und Betreuung durch nicht behinderte Personen erwächst (HALLSTEIN, 1993). Da nicht behinderte Kinder nur von einer sehr begrenzten Zahl von Personen abhängig sind, Kinder mit Behinderung jedoch aufgrund ihres erhöhten Bedarfs an Pflege, Therapie und anderen Formen von Unterstützung dagegen einer oftmals hohen Anzahl von Erwachsenen ausgesetzt sind, unter deren Kontrolle sie stehen, besteht hieraus die potentielle Möglichkeit, dass sich sexuelle Übergriffe ergeben. (vgl. SENN, 1993). Des Weiteren sind Menschen mit Behinderung oft intimen Pflegehandlungen ausgesetzt, ohne entscheiden zu können, wer diese an ihnen durchführt bzw. in welchem Ausmaß. Somit besteht so durchaus die Möglichkeit, dass Grenzen verschwimmen und gerade für jüngere Kinder nicht mehr klar erkennbar ist, wo Pflege aufhört und wo Missbrauch beginnt.

Damit eng verbunden ist ein vierter Faktor, die wirtschaftliche Abhängigkeit. Oftmals verfügen Menschen mit Behinderungen nicht über die notwendigen eigenen finanziellen und materiellen Mittel, um am wirtschaftlichen Leben teilnehmen zu können bzw. dürfen nicht eigenständig darüber entscheiden. Aus dieser andauernden Fremdbestimmung entsteht oft sehr angepasstes Verhalten, was in entsprechenden Macht- und Abhängigkeitsverhältnissen und potentielltem Risiko sexuellen Missbrauchs resultieren kann. So stellt SOBSEY (1994), dass Personen mit intellektuellen Defiziten irgendwann nicht mehr in der Lage sind, abzugrenzen, welche Anweisungen von Erwachsenen angemessen sind und welche nicht. So wird die geforderte Zustimmung gegenüber Autoritätspersonen möglicherweise auf alle Lebensbereiche übertragen.

Kommunikationsdefizite

Die verbale Handlungsfähigkeit kann bei Menschen mit Intelligenzminderung eingeschränkt sein, weshalb auch aufgrund kognitiver Defizite eventuelle Täterstrategien wie Manipulation, Bestechung und Drohungen oft nur unzureichend verstanden werden können. An die Opfer wird die Anforderung gestellt, in kürzester Zeit kognitive, verbale und körperliche Strategien zu entwickeln, die den Versuch eines Übergriffs vereiteln sollen (vgl. BECKER, 1995). Mit derart komplexen Prozessen sind Betroffene oftmals weit überfordert und somit wehrlos.

Noak und Schmid (1994) fanden, dass 17,2 % der befragten Mitarbeiter die eingeschränkte Artikulationsfähigkeit geistig behinderter Menschen als Risikofaktor für sexuellen Missbrauch angeben. Somit wirken sich Probleme in der Kommunikation doppelt negativ aus. Zum einen aufgrund der genannten situativen Überforderung, zum anderen aufgrund der Schwierigkeiten, anderen von sexuellen Gewalterfahrungen zu berichten. Behinderte Menschen verfügen oftmals nicht über das passende Vokabular, um die Übergriffe (adäquat oder nach Meinung oft unprofessioneller involvierter Personen) zu beschreiben bzw. werden nonverbale Signale falsch interpretiert oder erst gar nicht ernst genommen.

Die Wahrscheinlichkeit der Aufdeckung sexuellen Missbrauchs an geistig behinderten Menschen ist nach HALLSTEIN (1993) ohnehin gering.

Gesellschaftliche Normen und Einstellungen

Das Thema 'Sexualität geistig behinderter Menschen' scheint in der „modernen“ Gesellschaft noch immer einerseits verdrängt und geleugnet zu werden, andererseits werden behinderten Menschen vielfach nicht die gleichen Rechte und Ansprüche in Bezug auf Sexualität zugestanden wie nicht behinderten Menschen (vgl. NOAK & SCHMID, 1994). Diese 'Entsexualisierung' geistig behinderter Menschen steht in direktem Zusammenhang mit der fatalen Meinung oder Vorstellung, Menschen mit intellektuellen Einschränkungen würden ohnehin nicht begreifen, was mit ihnen passiert und somit auch nicht durch sexuelle Übergriffe psychisch oder in einer anderen Art und Weise davon beeinträchtigt werden (vgl. BECKER, 1995). Nach WALTER (1992) würden diese aufgrund ihrer Intelligenzdefizite und der mangelnden Differenzierungsfähigkeit ohnehin die Vorfälle oder Übergriffe schnell vergessen. Somit stellt sich die Negierung sexueller Gewalt zum Teil als gesellschaftliche Realität dar. Eine behinderte Frau entspricht oftmals nicht der gesellschaftlichen Norm einer attraktiven Frau. Der Irrglaube, dass sexuelle Attraktivität bei sexuellem Missbrauch eine Rolle spielt, bestimmt noch immer das Denken und lenkt davon ab, „dass es sich beim sexuellen Missbrauch um sexuelle Gewalt und nicht um gewaltsame Sexualität handelt“ (Hallstein, 1996, S.251, zitiert nach KLJAJIC, 2003).

Somit besteht die Gefahr unter Umständen einen potentiellen Täter dazu zu veranlassen, seine inneren Hemmungen gegen sexuelle Gewalt zu überwinden und sexuelle Übergriffe als unbedenklich oder belanglos einzustufen.

Distanzloses Verhalten

Unter distanzlosem Verhalten ist ein unangemessenes Interaktionsverhalten zu verstehen, bei dem der Betreffende auch „mit fremden Menschen unangemessen vertraulich, direkt oder sexuell enthemmt umgeht“ (MÖLLER, LAUX & DEISTER, 2001, S. 573). Distanzloses Verhalten ist häufiger bei Menschen mit einer geistigen Behinderung vorzufinden.

Oft werden sexuelle Handlungen geistig behinderter Menschen als bloße Befriedigung körperlicher Bedürfnisse angesehen, da sie aufgrund ihrer Behinderung nicht befähigt seien, ihre Sexualität in sozial akzeptierter Weise zu erfüllen. (vgl. WALTER, 1992). Nonverbale Kommunikationssignale vieler sprachunfähiger geistig behinderter Menschen werden fehlinterpretiert und als „klebrige Distanzlosigkeit fehlgedeutet und häufig sogar sexualisiert“ (WALTER, 1992, S.101).

Mangelnde oder geringe Glaubwürdigkeit

Oftmals wird Menschen, insbesondere jenen mit kognitiven oder sprachlichen Einschränkungen nicht geglaubt, wenn diese versuchen über einen erlebten sexuellen Missbrauch zu sprechen, was damit begründet wird, dass intellektuell beeinträchtigte Menschen als leicht verwirrbar gelten und die Bedeutungen von Begriffen, auf die sich sexuelle Handlungen beziehen, nicht verstehen.

Hard (1987, zitiert nach BECKER, 1995) fand heraus, dass 35 % der Frauen mit intellektueller Beeinträchtigung, die versuchten, anderen Menschen den Missbrauch zu berichten, nicht geglaubt wurde. Auch 25 % der Frauen, denen geglaubt wurde, wurden weiterhin sexuell missbraucht. Die Tatsache, dass behinderten Menschen oftmals nicht geglaubt wird, wiegt die Täter in Sicherheit, da sie davon ausgehen, dass sie nur mit geringer Wahrscheinlichkeit gefasst und verurteilt werden, was die Gefahr eines Übergriffs erhöhten kann. So stellte SOBSEY (1994) beispielsweise heraus, dass nur 24,3 % der Fälle sexuellen Missbrauchs gegen intellektuell beeinträchtigte Menschen überhaupt angezeigt und davon nur etwa ein Drittel der Täter tatsächlich verurteilt wurde. Hinzu kommt, dass in 32,1 % der Fälle eine Anzeige durch die Polizei, Staatsanwaltschaft oder durch das Gericht abgelehnt wurde. Auch ZEMP & PIRCHER (1996) bestätigen, dass nur wenige Berichte über erlebte sexuelle Gewalt behinderter Frauen Konsequenzen nach sich ziehen.

Bezieht man hier erneut das Modell von Finkelhor mit ein, so können mangelnde Konsequenzen und damit das Bewusstsein, einem „sicheren Opfer“ gegenüberzustehen, durchaus einen Faktor darstellen, die innere Hemmung, einen Menschen zu missbrauchen, zu überwinden.

Nach Finkelhor (1994) verfügen Kinder über kein hinreichendes soziales Netzwerk, das im Falle eines sexuellen Missbrauchs Unterstützung und Begleitung bietet und als Anlaufstelle dienen bzw. eingeschaltet werden kann (vgl. FINKELHOR, 1994) Dieses Wissen potentieller Täter könnte ebenfalls sexuelle Übergriffe begünstigen.

Mangelnde Information bezüglich Sexualerziehung

Erhöhtes Missbrauchsrisiko kann des Weiteren in Zusammenhang gebracht werden mit dem Versäumnis, behinderten Menschen angemessene Informationen über ihren Körper und

Sexualität zu vermitteln. Da Menschen mit Entwicklungsverzögerungen jedoch oftmals von Familie und Pflegepersonen jegliche Sexualität abgesprochen wird, mangelt es zum Teil an sexueller Aufklärung und Präventionsarbeit, was das Risiko erhöht, Opfer sexueller Übergriffe zu werden. (vgl. SENN, 1993). SOBSEY (1994) berichtet, dass einige junge Frauen aus der Stichprobe von Tätern vermittelt bekommen hatten, sie hätten die sexuellen Übergriffe als Teil eines Sexualerziehungsprogramms zu ertragen. Dass angemessene Sexualerziehung von behinderten Menschen noch immer ein Tabu darstellt, zeigt die Untersuchung von ZEMP & PIRCHER (1996). So wussten jeweils etwa nur die Hälfte der befragten Frauen etwas über Geschlechtsverkehr und sexualisierte Gewalt. Etwa ein Drittel der Frauen war sexuell nicht aufgeklärt.

Allerdings besteht durchaus eine Beziehung zwischen Sexualerziehung und sexuellem Missbrauch.

Hard (1987, zit. nach BECKER, 1995) kam zu dem Ergebnis, dass alle befragten Frauen, denen es an angemessener Sexualerziehung mangelte, Opfer sexueller Übergriffe wurden im Gegensatz zu 12 % der Frauen, die Kenntnisse über Sexualität aufwiesen. Hard berichtet außerdem, dass Frauen in den Interviews erzählten, ihr gesamtes Wissen über Sexualität aus der Missbrauchserfahrung bezogen zu haben.

Demgegenüber kam STROMSNESS (1993) allerdings zu dem Ergebnis, dass Sexualerziehung und Aufklärung Missbrauch nicht vorzubeugen schien, allerdings erhöhte sich die Zahl der Opfer, die über den sexuellen Missbrauch berichteten.

Mangelndes oder geringes Selbstwertgefühl

Untersuchungen haben ergeben, dass nicht behinderte Menschen zu einem nicht unerheblichen Prozentsatz (44 %) über ein Gefühl der Ablehnung gegenüber Menschen mit einer Behinderung berichten. Des Weiteren wird von einer sozialen Distanz zwischen Menschen mit einer Behinderung und ohne eine solche ausgegangen (vgl. BECKER, 1995). Aufgrund einer geringeren Wertschätzung und einem niedrigeren sozialen Status scheint das Risiko erhöht, Opfer eines sexuellen Übergriffs zu werden, „Bestimmte Einstellungen können dazu führen, dass individuelle Hemmschwellen, geistig behinderte Mädchen sexuell zu missbrauchen, herabgesetzt werden“ (BECKER, 1995, S. 99).

So entwickelt sich Identität, Selbstwert und Selbstbewusstsein in Abhängigkeit von der sozialen Umgebung. Anforderungen und Vertrauen in die kindlichen Fähigkeiten seitens der Eltern sind wichtige Voraussetzungen dafür, dass das Kind ein positives Selbstbild entwickelt (RAT für KRIMINALITÄTSVERHÜTUNG, 2003). Auch die Beziehungen zu Gleichaltrigen sind wichtig für die Ausbildung eines Selbstbewusstseins. Eine völlige Akzeptanz der Kinder mit Behinderung ist allerdings oftmals weder im Familienkreis noch bei gleichaltrigen, nicht behinderten Kindern gegeben. Kinder mit Behinderung leben oftmals mit dem Anspruch, ihre

Defizite auszugleichen, was Mädchen und Jungen schon früh zeigt, dass nicht die eigenen Ansprüche und Bedürfnisse vordergründig sind, sondern die der sozialen Umwelt. Auch beim Versuch, Ansprüche von außen zu erfüllen, werden eventuell eigene Bedürfnisse verdrängt, was die Entwicklung einer eigenen Identität erschwert (vgl. RAT für KRIMINALITÄTSVERHÜTUNG, 2003). Gerade ein starkes Selbstbewusstsein, die Vertretung des eigenen Standpunkts zur Berücksichtigung eigener Bedürfnisse, Selbstbestimmung und Durchsetzungsvermögen sind wichtige Kompetenzen zur Prävention sexuellen Missbrauchs. Die besonderen Sozialisationsbedingungen von Menschen mit Behinderung sind aber oftmals nicht auf diese speziellen Fähigkeiten ausgelegt, sondern vielmehr auf Anpassung, Akzeptanz und Zustimmung, wodurch wichtige Grundvoraussetzungen fehlen, um sich gegen sexuellen Missbrauch zu wehren.

In Sobsey's Untersuchung (1994) zeigte sich, dass 24,4 % der Hilfsstellen angaben, dass Menschen mit geistiger Behinderung ein extrem angepasstes Verhalten oder mangelnde Selbstsicherheit zeigten und in diesem Zusammenhang auch mit einem erhöhten Risiko lebten, Opfer sexueller Übergriffe zu werden.

Zusammenfassend lässt sich feststellen, dass Behinderung oder intellektuelle Beeinträchtigung per se keinen Bedingungs- oder Risikofaktor für sexuelle Gewalt darstellt, sondern dass vielmehr gesellschaftliche und soziale Aspekte sowie das Zusammenwirken der verschiedenen genannten Faktoren, welchen Menschen mit Behinderung unterliegen, einen Risikofaktor für sexuelle Gewalt darstellen.

1.3.2.3 Aspekte zum Täterverhalten

Die Täterinnen und Täter kommen aus dem gesamten Umfeld der Opfer, vor allem dem so genannten sozialen Nahraum. Hier sind vor allem der familiäre Bereich, institutionelle Einrichtungen, der ärztlich-therapeutische und schulisch-(sozial)pädagogische Bereich sowie der Freizeit- und Arbeitsbereich allgemein zu nennen. Die große Nähe zu den Pflegepersonen ermöglicht dem Täter ständige Möglichkeiten und Gelegenheiten zu Übergriffen und eventuellen Sanktionsmaßnahmen. Weiterhin ermöglichen bzw. geben die extreme Verletzbarkeit sowie Abhängigkeitsverhältnisse dem Täter leichten wiederholbaren Zugriff sowie die weitestgehende Sicherheit vor Veröffentlichung oder Aufdeckung des Missbrauchs vor. 97 % der Täter sind Männer. (vgl. RAT für KRIMINALITÄTSVERHÜTUNG SCHLESWIG-HOLSTEIN, 2003).

Neben den schon dargestellten Risikofaktoren postuliert Finkelhor (1984, 1986, zit. nach SENN, 1993) vier weitere Vorbedingungen, von denen sexueller Kindesmissbrauch abhängig ist, die vor allem im Zusammenhang mit dem Täterverhalten (Psyche des Täters) stehen:

- (1) Der Erwachsene muss prädisponiert sein für den Missbrauch (Motivation)
- (2) Innere Sperren oder Hemmungen gegen das Befolgen dieser Motivation bzw. gegen den Missbrauch müssen reduziert bzw. überwunden werden
- (3) Äußere (soziale) Hemmungen müssen abgebaut werden
- (4) Der Widerstand und die Schutzmechanismen des Opfers müssen überwunden werden, sei es durch den Täter selbst oder andere Faktoren (vor allem situative Bedingungen des Opfers)

Daneben nennen BROXHOLME & LINDSAY (2003) ergänzend einige Aspekte, wie etwa den Mangel an einer normalen Entwicklungsumwelt, an geeigneten Partnern, Privatsphäre, an sexueller Aufklärung, an juristischem Wissen, an generell sozialen und kognitiven Fähigkeiten. Dazu kommen oftmals eine verzögerte Sexualentwicklung, restriktive Umwelteinflüsse, Ab- und Ausgrenzungseffekte, irrealer sexuelle Vorstellungen und Erwartungen, Drogeneinflüsse und Ähnliches.

Die Tatsache der Behinderung des Kindes oder Jugendlichen kann somit dazu führen, den Täter im Sinne eines Prozesses der Dehumanisierung in seinen Vorstellungen zu bestärken, den Übergriff so zu entschuldigen oder zu rechtfertigen, indem die Person als ein (unmenschliches) Objekt betrachtet wird. Persönliche und soziale Hemmungen gegen Übergriffe könnten auf diese Weise reduziert werden (vgl. SENN, 1993).

Trotzdem zwar einige Forscher (vgl. SENN, 1993) der Ansicht sind, Menschen würden eher Mitleid für Kinder mit Behinderung empfinden und sie daher nicht als mögliche Opfer sehen, ergeben sich insgesamt und zusammenfassend kaum Argumente für die Unterstützung der These, dass eine Behinderung einen protektiven Faktor für sexuellen Missbrauch darstellt. So ist davon auszugehen, dass Kinder und Jugendliche mit kognitiven Defiziten einem erhöhten Risiko ausgesetzt sind, physisch und/oder psychisch misshandelt zu werden.

In den letzten Jahren erfolgte zunehmend mehr, vor allem neuropsychologische Forschung bezüglich der Gehirnstrukturen und -prozesse von Sexualstraftätern. Insbesondere erwiesen sich zwei Arten von Defekten als wichtig. Zum einen Defekte im Bereich des Kortex, die sich in Hemmungsreduktion zeigen, zum anderen mögliche Defekte im limbischen System, die sich in abartigen Sexualtrieben manifestieren können, vor allem in Form von intensiven Beziehungen zwischen Aggression und Sexualität. Dabei wird oft die aggressive Natur von Sexualverbrechen eher als Verbrechen im Rahmen von Aggression und Macht(-ausübung) interpretiert denn als reines Sexualverbrechen. Für viele Täter erscheint somit die Unterscheidung zwischen Aggression und Sexualität bedeutungslos, Macht und Sexualität werden

als 'Eins' gesehen. Die notwendige internale Hemmung und Selbstkontrolle ist dabei abhängig von externalen Hemmfaktoren bzw. deren Interaktion. Verschiedene Komponenten des persönlichen Microsystems als auch des Macrosystems spielen hierbei eine wichtige Rolle (vgl. BROXHOLME & LINDSAY, 2003; SENN, 1993).

Natürlich kann im Rahmen der Auseinandersetzung mit diesem Thema nicht von einer Reduzierung dieser Personengruppe allein auf den Opferstatus ausgegangen werden. Allerdings tauchen auch hier Menschen mit Behinderungen in Erhebungen weder als Opfer noch als Täter explizit auf. Dies betrifft vor allem den Bereich von Wohneinrichtungen und Werkstätten für behinderte Menschen, wo es Aussagen von Mitarbeitern und Bewohnern gibt, die Mitbewohner als Täter nennen. Dabei ergeben sich Probleme und Unsicherheit im Umgang Fragen der Glaubwürdigkeit, Schuldfähigkeit und eventueller Handlungsmöglichkeiten. Auch hier besteht demnach dringender Forschungsbedarf, um sich dem Thema schrittweise weiter anzunähern (vgl. RAT für KRIMINALITÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN, 2003).

1.3.2.4 Rechtslage und Strafrechtsnormen

Aus strafrechtlicher Sicht sind unter dem Titel 'Gewalt gegen Menschen mit Behinderung' mehr zu subsumieren als nur die typischen Gewaltdelikte (z. B. Vergewaltigung - § 177 StGB, Nötigung - § 240 StGB, ...), sondern vielmehr alle Straftaten, bei denen Menschen mit Behinderung unter einem Über- und Unterordnungsverhältnis leiden und körperlichem oder seelischem Zwang ausgesetzt sind; also alle Straftaten gegen die sexuelle Selbstbestimmung (§§ 174-184 StGB), wobei es keinen Unterschied macht, ob eine Behinderung vorliegt oder nicht (vgl. WALTER, 1992).

Für den Bereich der Straftaten gegen die sexuelle Selbstbestimmung im Zusammenhang mit Gewalt gegen Menschen mit Behinderung sind besonders § 174 StGB generell (sexueller Missbrauch von Schutzbefohlenen), § 174c StGB (sexueller Missbrauch unter Ausnutzung eines Beratungs-, Behandlungs- oder Betreuungsverhältnisses) und § 179 StGB (sexueller Missbrauch widerstandsunfähiger Personen) hervorzuheben.

Daneben gilt es auf die Gleichstellung von Menschen mit Behinderungen im Prozessrecht hinzuweisen, wonach diese im Rahmen eines Prozesses - egal auf welcher Seite sie stehen - nicht benachteiligt werden dürfen. Neben den gleichen Rechten haben sie jedoch auch die gleichen Pflichten wie nicht beeinträchtigte Personen.

So muss jedem Zeugen - gleich wie gearteter Behinderung oder Beeinträchtigung – die Möglichkeit gegeben werden, eine Zeugenaussage abzugeben, da jeder Mensch die Fähigkeit besitzt, Zeuge zu sein, unabhängig davon, ob eine geistige oder körperliche Beeinträch-

tigung vorliegt. Dazu müssen alle Möglichkeiten und Voraussetzungen geschaffen werden, sei es technischer Natur oder sonstiger Hilfsmittel und Unterstützungen (vgl. RAT für KRIMINALITÄTSVERHÜTUNG SCHLESWIG-HOLSTEIN ,2003).

1.4 Menschen mit Intelligenzminderung als Zeugen

In der juristischen Literatur (bezogen auf das Rechtssystem der Bundesrepublik, BENDER & NACK, 1995B, Randnr. 944 ff), wird darauf verwiesen, dass die Bewertung und Würdigung von Zeugenaussagen die ureigenste Aufgabe des Richters ist, dieser aber gehalten sei, in Fällen, in denen die Aussagetüchtigkeit eines Zeugen, sprich seine Wahrnehmungs- und Erinnerungsfähigkeit „über das übliche Maß hinaus beeinträchtigt“ sei, entsprechende (meist psychologische) Sachverständigengutachten einzuholen.

Wenn intellektuell beeinträchtigte Zeugen vor Gericht aussagen, wird ihnen oftmals aus verschiedenen noch näher darzustellenden Gründen eine verminderte Glaubwürdigkeit zugesprochen, was eine Untersuchung von PELED et al. (2004) eindrucksvoll bestätigt. Dabei sollten Versuchspersonen anhand von Zeugenaussagen und Berichten von Augenzeugen allgemeine Fragen zur Glaubwürdigkeit der Schilderungen beantworten und diese einschätzen. Diejenigen Versuchspersonen, die wussten, dass und welche Aussagen von lernbehinderten 15-jährigen Jugendlichen stammten, schätzten diese deutlich unglaubwürdiger ein als die einer Kontrollgruppe von normal begabten 10-jährigen Kindern. War diese Information nicht bekannt, fanden sich keine signifikanten Unterschiede (vgl. PELED et al., 2004).

Demgegenüber fanden STOBBS & KEBBEL (2003), dass lernbehinderte Zeugen zwar als genauso wahrheitsliebend und ehrlich wahrgenommen und eingeschätzt werden wie Personen der übrigen Bevölkerung, jedoch wurden Augenzeugenberichte letzterer trotzdem für glaubhafter, korrekter, genauer und höherwertig (im Sinne einer als „guter“ Zeuge wahrgenommenen Person) sowie die Personen selbst für kompetenter befunden. Da es ja tatsächlich so ist, dass Menschen mit intellektuellen Einschränkungen oft unterschiedliche Schwierigkeiten beim Enkodieren, Speichern und Behalten von Informationen und Erinnerungen haben und diese demnach auch nicht als homogene Gruppe angesehen und behandelt werden sollten, ist im Rahmen einer Zeugenaussage oder Befragung eine Einschätzung der individuellen Fähigkeiten erforderlich und in jedem Fall angebrachter als sich auf globale Eindrücke und Stereotype zu verlassen (vgl. STOBBS & KEBBEL, 2003).

Es erscheint daher sowohl aus Sicht der psychologischen Forschung, aber auch aus juristischer Sicht angeraten, sich verstärkt mit den Kompetenzmerkmalen dieser Personengruppe auseinander zu setzen.

1.4.1 Beurteilung der Aussagetüchtigkeit bei Menschen mit Intelligenzminderung

Nach GREUEL et al. (1997) ist die Einschätzung individueller Charakteristika des zu begutachtenden Zeugen ebenso von Bedeutung. In diesem Zusammenhang stellt sich die grundlegende Frage, ob ein Zeuge über die intellektuellen und kognitiven Grundvoraussetzungen verfügt, um eine vor Gericht zu verwertende Aussage abzugeben. Andererseits muss auch abgewogen und bedacht werden, ob ein Zeuge über die kognitiven Grundvoraussetzungen verfügt, eine (intentionale) Falschaussage zu generieren und gegebenenfalls über mehrere Befragungszeitpunkte aufrechtzuerhalten (vgl. GREUEL et al., 1997).

Allgemein kann nach GREUEL et al. (1998, S.79) unter der *Aussagetüchtigkeit oder Aussagefähigkeit* die Fähigkeiten eines Zeugen verstanden werden,

- „den in Frage stehenden Sachverhalt zuverlässig wahrzunehmen,
- ihn in der zwischen dem Geschehen und der Befragung liegenden Zeit im Gedächtnis zu bewahren,
- über ausreichendes Sprachverständnis für die Befragung sowie über ausreichende sprachliche Ausdrucksfähigkeit für die Schilderung des Geschehnisses zu verfügen,
- ein ausreichendes Maß an Kontrollmöglichkeiten gegenüber Suggestiveinflüssen zur Verfügung zu haben sowie
- Erlebtes von Phantasievorstellungen unterscheiden zu können“.

Allerdings ist hierbei abzugrenzen, dass unter der Zeugentüchtigkeit kein situationsübergreifendes personales Merkmal verstanden wird. Für die Beurteilung der Aussagefähigkeit ist die Ausprägung der Funktionen, die während des in Frage stehenden Sachverhalts die Wahrnehmung, in der zwischen Erlebnis und Aussage liegenden Zeit das Gedächtnis und während der Aussage das Erinnern beeinflusst haben könnten, bedeutsam (vgl. GREUEL et al., 1998).

In vielen Fällen wurde gerade Zeugen mit Intelligenzminderung eine allgemeine Eignung als Zeugen mit der Begründung abgesprochen, dass es ihnen an der Kompetenz mangle, Erlebtes von Phantasiertem zu unterscheiden und sie gegenüber Suggestionseinflüssen besonders anfällig seien. Ebenfalls wurde angenommen, dass geistig behinderte Zeugen nicht über die Fähigkeit verfügen, ein erlebtes Ereignis genau wiederzugeben (vgl. DENT, 1986) und ihre Gedächtnissysteme angeborenerweise defekt seien (vgl. PERLMAN et al., 1994; BEAIL, 2002). Da, wie bereits dargestellt, Menschen mit intellektueller Beeinträchtigung nicht selten und immer öfter Opfer von sexueller Gewalt werden, ist eine grundsätzliche Absage an die Zeugeneignung dieser Personengruppe problematisch und nicht ungefährlich.

Eine reine Orientierung am Intelligenzniveau erscheint bei der Beantwortung der Frage, ob ein Zeuge in der Lage ist, einen Sachverhalt zuverlässig wahrzunehmen und wiederzugeben, recht unzureichend, da es nach GREUEL et al. (1998) kein festgelegtes Intelligenzniveau gibt, das für die Eignung als Zeuge erforderlich ist. Somit seien – zwar oftmals eingeschränkt, aber grundsätzlich möglich – auch Personen mit verminderten intellektuellen Fähigkeiten zu genauer Beobachtung und wirklichkeitsgetreuem Behalten im Gedächtnis, der Rekonstruktion und Wiedergabe in der Lage. Somit muss in diesem Kontext die Beurteilung der Aussagetüchtigkeit immer individuell erfolgen und sollte keinesfalls pauschal am grundsätzlichen Vorliegen einer Behinderung oder sonstigen Beeinträchtigung festgemacht werden.

Weiterhin ist es wichtig zu erwähnen, dass die Kompetenzfrage durchaus situationsspezifisch zu beurteilen ist. So können minderbegabte Zeugen beispielsweise Schwierigkeiten haben, einen als Augenzeuge erlebten, komplexen Verkehrsunfall zu schildern, andererseits aber durchaus in der Lage sein, autobiographische Erinnerungen korrekt abzurufen und zu schildern oder glaubhafte Aussagen über ein selbst erlebtes Delikt gegen die sexuelle Selbstbestimmung zu produzieren (vgl. MÜLLER-LUCKMANN, 1980).

1.4.1.1 Aspekte der Zeugenkompetenz von Menschen mit kognitiven Einschränkungen

Nach MEYER-GOßNER (1997) gibt es im juristischen Sinn keine allgemeine Zeugnisunfähigkeit. Zeuge kann demnach auch sein, „wer körperliche oder geistige Gebrechen hat, sofern er nur zu Wahrnehmungen und ihrer Wiedergabe vor Gericht fähig ist“ (vgl. MEYER-GOßNER, 1997). In Fällen von zum Teil schweren und mehrfachen Behinderungen ist eine Aussageunfähigkeit jedoch wohl gegeben, da Wahrnehmung, Gedächtnis und verbale Fähigkeiten zu erheblich beeinträchtigt sind.

Wie bereits beschrieben, wird zwar schon seit mehreren Jahren gefordert, das in der allgemeinen und in der Psychologie der Zeugenaussage vorhandene Forschungsdefizit hinsichtlich der Zeugenkompetenz und der Bewertung von Aussagen von Personen mit Intelligenzdefiziten auszugleichen (PERLMAN et al., 1994; BULL, 1995A). Dennoch ist die diesbezügliche Forschungslage bislang noch sehr überschaubar, es finden sich nur einige wenige Untersuchungen, die sich gezielt mit der genannten Thematik auseinandersetzen.

Ein mehrfach untersuchter Komplex betrifft die Zeugenkompetenz und die Befragung von Personen mit Lernbehinderung bzw. leichter geistiger Behinderung (DENT, 1986; BROWN & GEISELMAN, 1990; DENT, 1992; CLARE & GUDJONSSON, 1993; PERLMAN et al., 1994;

BULL, 1995c, CARDONE & DENT, 1996; MILNE, 1999; MILNE, CLARE & BULL, 1999; GUDJONSSON, MURPHY & CLARE, 2000; MILNE & BULL, 2001; AGNEW & POWELL, 2004; HENRY & GUDJONSSON, 2004).

Die Ergebnisse der genannten Untersuchungen zeigen weitgehend übereinstimmend, dass freie Berichte lernbehinderter Personen oft zwar weniger vollständig ausfallen als Berichte einer gesunden Vergleichsstichprobe, was vermutlich auf eine generell verlangsamte Informationsverarbeitung zurückgeht (MILNE & BULL, 2001), im Verhältnis aber nicht mehr fehlerhafte Details (DENT, 1992; BULL, 1995; MILNE, CLARE & BULL, 1999), sondern auch hochspezifische ereignisbezogene Details (AGNEW & POWELL, 2004) enthalten. Eine verstärkte Tendenz zu Intrusionen oder Konfabulationen in dem Sinn, dass zuvor nicht präsentierte Information Eingang in die Berichte der Versuchspersonen findet, ließ sich bei lernbehinderten Versuchspersonen bisher nicht nachweisen. Insgesamt fanden sich keine signifikanten Korrelationen zwischen intellektueller Fähigkeit, verbalem Erinnerungsgedächtnis, Suggestibilität oder Zustimmungstendenz (GUDJONSSON & CLARE, 1995). Entscheidend hierbei ist offensichtlich, ob die Personen über einzelne Vorfälle oder verschiedene, mehrere Vorfälle eines wiederholten Ereignisses berichten (POWELL & THOMSON, 2003). Das größte Problem liegt nach Auffassung von MILNE & BULL (2001) eher im Abruf und der Wiedergabe der gespeicherten Informationen, was durch Schwierigkeiten in Sprachverständnis und –ausdruck noch verstärkt wird.

Untersuchungen zur Befragung Lernbehinderter zeigen, dass deren Erinnerungsleistung durch Verwendung entsprechender Gedächtnishilfen wie dem Kognitiven Interview (GEISELMAN & FISHER, 1992) in ähnlicher Weise wie bei normal begabten Versuchspersonen gesteigert werden kann (BROWN & GEISELMAN, 1990; MILNE, CLARE & BULL, 1999; MILNE & BULL, 2001), das heißt, dass freie Berichte der Versuchspersonen deutlich mehr korrekt wiedergegebene Details enthalten als bei Anwendung eines strukturierten Interviews ohne Gedächtnishilfen, und dass die Berichte im Verhältnis aber nicht mehr, sondern tendenziell weniger fehlerhafte Details enthalten.

Problematischer gestaltet sich das Antwortverhalten von Versuchspersonen mit Lernbehinderung auf Fragen. Hier ließen sich zum Teil deutliche Antworttendenzen nachweisen. So wird z.B. den Ergebnissen von MILNE & BULL (2001) zufolge bei Alternativfragen häufiger die zweite Alternative gewählt; ebenso besteht offenbar die Tendenz, geschlossene Fragen eher (auch falsch) zu beantworten, als Erinnerungslücken einzuräumen. PERLMAN et al. (1994) kommen wie MILNE et al. (1999) zu dem Schluss, dass auch die Antworten auf offene Fragen hin deutlich unzuverlässiger seien als die im freien Bericht erfolgten Schilderungen.

HENRY & GUDJONSSON (2004) verglichen Kinder im Alter zwischen 10 und 13 Jahren mit leichten (IQ 55-70) und mittleren (IQ 40-54) intellektuellen Einschränkungen mit normal begabten Gleichaltrigen anhand einer Augenzeugengedächtnisaufgabe. Dabei wurde die Stärke der Gedächtnisspuren verändert, um zu untersuchen, ob eine Erhöhung der Stärke der Gedächtnisspur eher für die intelligenzgeminderten oder die normal begabten Kinder von Vorteil sei. Dies konnte ebenso wenig bestätigt werden, wie die Annahme, dass in den beiden Gruppen Gedächtnisprozesse unterschiedliche Mechanismen zur Folge haben.

In mehreren Untersuchungen (CARDONE & DENT, 1996; GUDJONSSON et al., 2000; MILNE & BULL, 2001) haben sich Versuchspersonen mit Lernbehinderung als deutlich anfälliger für Suggestivfragen bzw. den Falschinformationseffekt erwiesen als normal begabte Versuchspersonen. Die von GUDJONSSON (1992) oder CECI & BRUCK (1995) angeführten Erklärungsmodelle für die Wirkung suggestiver Beeinflussung (Compliance und Antworttendenzen gegenüber dem Interviewer, Akzeptanz von Autoritätspersonen, Konzentrationsmängel, mangelnder Wortschatz) zeigen sich nach Auffassung der zitierten Autoren bei lernbehinderten Versuchspersonen verstärkt. In neueren theoretischen Ansätzen wird Suggestibilität bzw. der Falschinformationseffekt als Problem der Quellenidentifikation bzw. der Realitätsüberwachung verstanden (LINDSAY, 1994; HIGHAM, 1998).

Ganz grundsätzlich ist die Fähigkeit, unterschiedliche Informationsquellen zu diskriminieren, für Zeugenaussagen zentral, da häufig nicht nur zwischen verschiedenen Fremdquellen (z.B. unterschiedlichen Tatorten), sondern auch zwischen internalen und externalen Quellen unterschieden werden muss. Aschermann (2002) hat diesbezüglich ein Modell der Quellenidentifikation aufgestellt, das diese unterschiedlichen Quellendimensionen abdeckt und damit die bislang vorherrschende Unterscheidung zwischen 'reality monitoring' bzw. 'internalem source monitoring' (Unterscheidung zwischen externalen und internalen Quellen) und 'externalem source monitoring' (Unterscheidung zwischen externen Quellen) integriert.

Um Aufschlüsse über die Zeugentüchtigkeit von Personen mit Lernbehinderung gewinnen zu können, erscheint dieser Punkt besonders essentiell. Während die zuvor zitierten Studien zwar gezeigt haben, dass im Verhältnis nicht mehr Intrusionen oder Konfabulationen in den Berichten lernbehinderter Versuchspersonen auftreten, erscheint die für die Zeugeneignung eigentlich wichtigere Frage, inwieweit Personen mit Lernbehinderung in der Lage sind, internal generierte von external erhaltenen Informationen abzugrenzen bzw. unterschiedliche externe Quellen zu differenzieren. Im Bereich der Forschung zur Quellenidentifikation wurden in bislang zwei Studien lernbehinderte Kinder untersucht. Im typischen experimentellen Quellenidentifikations-Paradigma werden üblicherweise Stimuli aus zwei oder mehr Quellen präsentiert (z. B. zwei Wortlisten), in der Abfrage werden diese Items mit Distraktoriems gemischt, die Versuchspersonen müssen für jedes Item zunächst eine 'alt/neu – Unterscheidung' treffen und dann die als „alt“ klassifizierten Items einer der Quellen zuord-

nen (vgl. BATCHELDER & RIEFER, 1990). LORSBACH, MELENDEZ & CARROLL-MAHER (1991) und LORSBACH & EWING (1995) stellten dazu fest, dass lernbehinderte Kinder, die an Quellenidentifikationsexperimenten teilnahmen, sowohl im Wiedererkennensteil, aber auch im Quellenidentifikations-Teil deutlich schlechter abschneiden als normalbegabte Kinder. Dies mag zum Großteil daran liegen, dass die oft komplexen Kontextinformationen entweder nicht spontan enkodiert wurden und/oder nicht systematisch abgerufen werden können (LORSBACH, 2000). In beiden Bedingungen ('internal source monitoring' und 'reality monitoring') sind verschiedene Gedächtnisleistungen und Entscheidungsprozesse gefordert respektive der Einordnung und Verarbeitung sowohl sensorischer Attribute als auch damit verbundener kognitiver Operationen.

LORSBACH (2000) schreibt, dass Gedächtnisprobleme lernbehinderter Kinder zwar nicht verallgemeinert werden können, jedoch scheinbar auf Aufgaben beschränkt sind, die sich auf bewusste Erinnerungsversuche beziehen. Ähnlich wie bei jüngeren Kindern haben diese Schwierigkeiten bei kognitiv anspruchsvollen und mühevollen sowie prozesskontrollierten Aufgaben, allerdings nicht bei automatisch auszuführenden Aufgaben oder Entscheidungsprozessen. Hierbei zeigten sich nach neuesten Erkenntnissen nur im Fall expliziter Gedächtnistestung Unterschiede zu normal begabten Kindern, nicht wenn die Testung implizit verläuft (vgl. LORSBACH, 2000).

Im Rahmen ihrer Untersuchung zur Gedächtnisleistung von lernbehinderten Schülern für alltägliche Informationen verweisen McNAMARA & WONG (2003) zunächst auf die *Hypothese der Zugänglichkeit (accessibility hypothesis)* von Informationen bzw. Erinnerung, die davon ausgeht, dass auch lernbehinderte Schüler zwar über die zur Lösung einer Aufgaben notwendigen Informationen verfügen, jedoch weniger in der Lage sind, darauf zuzugreifen. Inwieweit die genannte Hypothese auf nicht-experimentelle Bereiche oder Alltagsinformationen anwendbar ist, ist aufgrund der mangelnden Befunde aus der Forschung bezüglich lernbehinderter Menschen zunächst noch unklar. Von Bedeutung hierbei ist ebenfalls eine bestimmte Wissensform, das '*general event knowledge*' (auch generic personal memory) als Teil des autobiographischen Gedächtnisses. Das '*general event knowledge*' bezieht sich dabei auf allgemeine Wissensstrukturen von Ereignis-Prototypen, die aus der Erfahrung wiederholter Ereignisse hergeleitet sind (vgl. McNAMARA & WONG, 2003).

Das Arbeitsgedächtnis als dynamisches, aktives und zentral ausführendes Gedächtnissystem (BADDELEY, 1986) besteht dabei aus zwei passiven, in ihrer Kapazität beschränkten Subsystemen; dem für die temporäre Speicherung verbaler Information zuständigen 'phonologischen loop-system' sowie dem 'visuospatial-sketchpad-subsystem', das vor allem visuelle und räumliche Informationen speichert. Um nun alltägliche Erfahrungen zu enkodieren und im Anschluss zu erinnern, sind Entscheidungsprozesse wie auch interne Problemlösungen, logisches Denken und Überlegung erforderlich. Nach Swanson (1993, 1999; zit.

nach McNAMARA & WONG, 2003) haben lernbehinderte Schüler möglicherweise Defizite in der ausführenden Komponente des Arbeitsgedächtnisses, die allerdings unabhängig von irgendwelchen sprachspezifischen Verarbeitungsproblemen sind. Schlechte Erinnerungsleistungen könnten demnach entweder auf das Fehlen einer passenden Gedächtnisspur (*storage deficiency*) oder Schwierigkeiten beim Zugang während des Behaltensvorgangs (*retrieval deficiency*) zurückzuführen sein. Eine derartige Unterscheidung ist jedoch problematisch so zu treffen, da es methodisch schwierig ist, Speicherungs- bzw. Enkodierungsprozesse von Behaltensprozessen zu trennen. Swanson (1984; zit. nach McNAMARA & WONG, 2003) fand, dass lernbehinderte Schüler beim Erinnern nonverbaler, komplexer visueller Informationen bessere Leistungen erzielten als eine Kontrollgruppe. Die Untersuchung von McNAMARA & WONG (2003) selbst ergab, dass verglichen mit normal begabten Schülern, solche mit einer Lernbehinderung sowohl bei Aufgaben aus dem akademischen wie auch aus dem Bereich der alltäglichen Erinnerung schlechter abschnitten. Die Befunde deuten auf die ausgeführten Erklärungsmodelle hinsichtlich von fehlerhaften Prozessen des Arbeitsgedächtnisses sowie dem ungenügenden und nicht effektiven Gebrauch von Behaltensstrategien ebenso wie eines Abrufdefizits bezüglich bereits früher enkodierter Information hin.

Verschiedene Untersuchungen haben gezeigt, dass im Gegensatz zum impliziten Gedächtnis das explizite Gedächtnis bei Kindern mit intellektuellen Defiziten beeinträchtigt ist (Lorsbach & Worman, 1989; Lorsbach et al., 1992; Lorsbach & Worman, 1990; zit. nach LORSBACH, 2000). Aufgrund der bewussten und wohlüberlegten Art expliziter Gedächtnisaufgaben verlassen sich diese Menschen stark auf eher reflektierende Aktivitäten (wie Wiederholung, Wiederfinden und Finden von Gemeinsamkeiten). Dieser Gebrauch von reflektiven Prozessen gestaltet sich allerdings kritisch und problematisch, insbesondere bei Aufgaben der Quellenidentifikation. Die bekannten Forschungsergebnisse weisen allerdings darauf hin, dass gerade diese reflektiven Prozesse, die entscheidend für den Aufbau differenzierter (perzeptueller wie semantischer) Informationen und kognitiver Operationen sind, bei intelligenzgeminderten Kindern gestört sind. Davon unbeeinträchtigt sind Gedächtnisleistungen und Aufgaben aus dem impliziten Bereich, da hierbei die Verarbeitung undifferenzierter Information gefordert ist.

KRINSKY-McHALE et al. (2003) untersuchten Alterseffekte und IQ-Effekte in Bezug auf implizites Gedächtnis bei Erwachsenen mit leichter und mittlerer geistiger Behinderung mit und ohne Down-Syndrom. Im Ergebnis zeigten sich jeweils signifikante, aber geringe Altersunterschiede bzw. IQ-bezogene Effekte. Ebenso beschäftigten sich ATWELL et al. (2003) mit intelligenz-bezogenen Unterschieden bei explizitem und implizitem Lernen von geistig behinderten Jugendlichen und jungen Erwachsenen. Dabei erzielten die Versuchspersonen mit geistiger Behinderung schlechtere Leistungen beim expliziten Lernen als die normal begabte Kontrollgruppe, jedoch vergleichbar gute bei impliziten Lerninhalten. Die Ergebnisse

deuten darauf hin, dass Lernen von komplexen Sachverhalten bei Personen mit und ohne geistige Behinderung funktional äquivalent ist, wenn dies durch implizite Verarbeitung ausgeführt wird (vgl. ATWELL et al, 2003).

Auch Perner et al. (1999; zit. nach ROEBERS & SCHNEIDER, 2001A) weisen darauf hin, dass der Einfluss von Intelligenz auf die Gedächtnisleistung auch altersabhängig zu sein scheint. Während dem chronologischen Alter bei jüngeren Kindern eher eine kritischere Rolle als der Intelligenz zukommt, nehmen bei älteren Kindern die Korrelationen zwischen Intelligenz und Erinnerungsleistungen zu.

Abseits dieser auf der Basis experimenteller Untersuchungen gewonnenen Erkenntnisse finden sich in der forensisch-psychologischen Literatur durchaus weitere Hinweise für den Umgang mit minderbegabten Zeugen. So schreiben GREUEL et al. (1998) unter Bezugnahme auf DETTENBORN (1989), es sei davon auszugehen, dass die Gedächtnisleistung minderbegabter Zeugen dahingehend beeinträchtigt sei, dass weniger das Gesamtgeschehen Gegenstand der Aussage ist, sondern eher ein Festhalten von Einzelinformationen erfolgt, was dazu jedoch eher zu einem Auslassen als zu einem Hinzufügen einzelner Erlebnisaspekte führe. Gerade die „geringer ausgebildete Fähigkeit zu gedanklicher Weiterverarbeitung und produktiver Phantasietätigkeit“ (S. 82/83) trage bei sachgerechter Befragung dazu bei, dass dieser Personenkreis weniger zu Verfälschungen, Ergänzungen oder Veränderung von Erlebnissen neige.

Allerdings trifft der Umstand, dass es bei Zeugenaussagen eher zu Auslassungen als zu Intrusionen kommt, auf nahezu jede Zeugenaussage zu und ist damit kein besonderes Merkmal der Aussagen Minderbegabter. Die von GREUEL et al. geäußerte Annahme, dass Personen mit kognitiven Einschränkungen weniger zu Aussageverfälschungen tendieren, impliziert, dass es sich hier quasi um in Teilbereichen zuverlässigere Zeugenaussagen handeln müsste als dies bei Normalbegabten der Fall ist. In Ermangelung einer nachvollziehbaren empirischen Fundierung besitzen diese Ausführungen aus unserer Sicht eher hypothetischen Charakter und sollten deshalb weiter überprüft werden.

1.4.1.2 Gedächtnisleistungen - Quellendifferenzierung und Realitätsüberwachung bei Intelligenzminderung

Bei Studien zur Gedächtniskapazität von Kindern und Jugendlichen mit Lernbehinderung sind diese nach Aspekten der strukturellen und funktionellen Kapazität zu gliedern.

Von Interesse bezüglich der strukturellen wie auch der funktionellen Kapazität sind vor allem der Vergleich mit Normalbegabten sowie der Einfluss des Stimulusmaterials und des Alters der Versuchspersonen.

Bei Betrachtung der inhaltsspezifischen Struktur des Langzeitgedächtnisses (vgl. BADDELEY, 2002), stellt sich zuerst einmal die Frage, ob kognitive Einschränkungen alle Komponenten des Langzeitgedächtnisses betreffen oder nur manche.

Die von Detterman (1979) formulierte Hypothese eines *'everything deficit'*, wonach Menschen mit kognitiven Einschränkungen in jeder Gedächtnisaufgabe schlecht abschneiden und daher eine Untersuchung von relativen Stärken und Schwächen ein sinnloses Unterfangen darstellt, kann aufgrund von gegenteiligen neuropsychologischen Forschungsergebnissen an amnestischen Patienten nicht aufrechterhalten werden und somit auch durchaus ein differenziertes Verhältnis aus Kompetenzen und Schwächen beobachtet werden. So werden beispielsweise bei Priming-Aufgaben, implizit erfragter Information oder Aufgaben, die die intentionale Wiedergabe von inzidentell oder/und automatisch enkodierter Information erfordern, bei Menschen mit Intelligenzminderung Leistungsniveaus der normal begabten Bevölkerung erzielt (vgl. VICARI & CARLESIMO, 2002).

Weiterhin untersuchten BURAK & ZIEGLER (1987) die Leistung von kognitiv eingeschränkten Menschen bei episodischen Gedächtnisaufgaben und variierten zwischen intentionalem und inzidentellem Stimuluslernen. Bei der freien Wiedergabe einer intentional gelernten Wortliste schnitten normalbegabte Kinder besser ab als Kinder mit kognitiven Defiziten. Wenn allerdings keine Lerninstruktion gegeben wurde, also inzidentelles Lernen und automatisches Enkodieren gefordert war, schnitten kognitiv beeinträchtigte Kinder ebenso gut ab wie die Kontrollgruppe. Zudem liegen der Wiedergabe von Zeugenbeobachtungen meist inzidentelle Beobachtungen zugrunde, da das Geschehen oftmals erst durch Vernehmung oder aussagepsychologische Exploration im Nachhinein mit besonderer Aufmerksamkeit belegt wird (vgl. GREUEL et al., 1998).

Von Relevanz ist in diesem Zusammenhang auch das von ARNTZEN (1993) postulierte *'Inkadenzphänomen'*, wonach auch normal begabte Menschen nicht immer und unter allen Umständen in der Lage sind, auf Informationen zuzugreifen. Somit müssen Erklärungsmodelle für Vergessen mit der mangelnden Fähigkeit, auf die Information zugreifen zu können, ergänzt werden. Nach MILNE & BULL (2001) kann diese Information durch angemessene Hinweisreize und passende Interviewstrategien jedoch wieder zugänglich gemacht werden.

Nach BEAIL (2002) kann ein schlechtes Abschneiden beim Abrufen von Erinnerungen zwar für kognitive Defizite im semantischen Gedächtnis sprechen, nicht aber unbedingt die Leistung vorhersagen, wenn ein Abruf aus dem autobiographischen Gedächtnis erfolgt. BEAIL (2002) führt weiter an, dass ein direktes Erleben Input von mehreren sensorischen Enkodierungssystemen beinhaltet und somit die Eingabe von mehreren Modalitäten zu einer elabo-

rierteren Gedächtnisspur im assoziativen Gedächtnis führt (vgl. BEAIL, 2002). So konnten CARDONE & DENT (1996) beispielsweise zeigen, dass eine multimodale Präsentation von Informationen eine vollständigere und genauere Erinnerung dieser Inhalte zur Folge hat. Lorschbach et al. (1992, zit. nach LORSBACH, 2000) untersuchte eventuell vorhandene Unterschiede hinsichtlich präsentierter Stimuli in Form von Bildern bzw. gesprochenen Worten sowie der Trennung von implizitem und explizitem Gedächtnis bei lernbehinderten und normalbegabten Kindern. Zwar war die Genauigkeit der Wiedererkennensleistung bei den Lernbehinderten geringer als in der Kontrollgruppe, jedoch ergaben sich keine Unterschiede bezüglich der Input-Modalität. Die Befunde decken sich mit der Überzeugung und der allgemeinen Ansicht, dass auf bildhaften Repräsentationen basierende Erinnerungen zugänglicher und leichter abzurufen sind als solche, die auf auditorischen oder verbalen Repräsentationen beruhen (*picture superiority effect*). Somit können verbale Quellenverwechslungen bei lernbehinderten Kindern verhindert werden, wenn sich die Quellen in hohem Maß unterscheiden. Insgesamt hängt eine problemlose und detaillierte Quellenzuordnung von der Menge und Qualität der ursprünglich wahrgenommenen Information ab, wobei sich Qualität diesbezüglich auf den Grad der Unterscheidbarkeit der Informationen bezieht. Dabei sind sowohl perzeptuelle als auch kontextuelle Einzelheiten von Bedeutung. Die bei lernbehinderten Personen gut dokumentierten Enkodierungsprobleme, also der mangelnden Fähigkeit komplexere und in hohem Maße differenzierbare Gedächtnisrepräsentationen zu bilden, resultieren somit möglicherweise in einer der Person zur Verfügung stehenden verminderten Menge und Spezifizierung der Information über die Quelle der unterschiedlichen Gedächtnisinhalte. Gedächtnismäßige Wahrnehmungen von Lernbehinderten enthalten demnach weniger differenzierte Informationen beispielsweise über den Klang unterschiedlicher Stimmen, kontextuell-semantiche Einzelheiten von Worten und den während des ursprünglichen Lernens verwendeten kognitiven Operationen. Wenn jedoch andererseits verschiedene Ereignisse präsentiert werden, die einzigartige, kennzeichnende Merkmale aufweisen (z.B. Bilder gegenüber gesprochenen Worten), sind die Gedächtnisinhalte auch für lernbehinderte Kinder leichter zu unterscheiden, was sich auch in zuverlässigeren und genaueren Quellenzuordnungen manifestiert (vgl. LORSBACH, 2000).

In der schon erwähnten Studie von GUDJONSSON & CLARE (1995) sollte der mögliche Zusammenhang zwischen (geringer) intellektueller Fähigkeit und der Tendenz zu Konfabulationen bei lernbehinderten Menschen im Alter von ca. 31 Jahren untersucht werden. Die zu testende Hypothese ging von einer positiven Korrelation aus. Die eingangs formulierte Hypothese konnte jedoch nicht bestätigt werden (vgl. GUDJONSSON & CLARE, 1995).

In Strafverfahren ist es immer wieder erforderlich, eine Personenidentifikation durch Augenzeugen (mit und ohne Intelligenzminderung) anhand eines Fotos vorzunehmen. Im Rahmen

ihrer Untersuchung zur Zuverlässigkeit bei Identifikationen (ERICSON & ISAACS, 2003) sahen Versuchspersonen (Erwachsene mit einer leichten Intelligenzminderung sowie eine Kontrollgruppe Normalbegabter) Filme über einen gestellten Diebstahl und hatten anschließend die Aufgabe, den Dieb in Fotoserien zu identifizieren. Es zeigte sich, dass die korrekte Trefferquote von Personen mit einer Intelligenzminderung genauso hoch war wie die in den Kontrollgruppen. Allerdings konnte im Vergleich zu den Menschen ohne eine Intelligenzminderung eine größere Tendenz zu raten, sowie eine erhöhte Quote von 'false alarms' festgestellt werden. Die Autoren weisen darauf hin, dass Falschidentifikationen nicht unbedingt problematisch bei Menschen mit Intelligenzminderungen sein müssen, da in den meisten Fällen eines sexuellen Missbrauchs die Täter dem Opfer wohl bekannt sind, da sie aus dem unmittelbaren sozialen Umfeld entstammen. Somit kann als eine wichtige Erkenntnis dieser Untersuchung festgehalten werden, dass die gefundenen Unterschiede nicht unbedingt auf eine schlechtere Gedächtnistätigkeit zurückzuführen sind, sondern eher auf soziale Fähigkeiten, die bei Entscheidungsprozessen erforderlich sind, wie beispielsweise das Verstehen der Aufgabenstellung und/oder der Umgang mit Autoritätspersonen (vgl. ERICSON & ISAACS, 2003).

Die Studie von GORDON et al. (1994) deckt sowohl den Bereich der allgemeinen Erinnerungs- und Gedächtnisleistung hinsichtlich des Konzepts der Realitätsüberwachung ab als auch die Frage, wie die Genauigkeit des Antwortverhaltens auf verschiedene Fragenformate variiert. GORDON und Kollegen verglichen in ihrer Untersuchung Kinder mit geistiger Behinderung mit durchschnittlich intelligenten Kindern gleichen Entwicklungsalters. Sie beschäftigten sich mit der Frage, inwieweit beide Gruppen sofort und nach einer sechswöchigen Verzögerung in der Lage sind, Aktivitäten, die tatsächlich passiert sind, von solchen zu unterscheiden, die sich die Kinder vorgestellt hatten. Zuerst wurde gemäß der Trichtertechnik ein freier Bericht angeregt, im Anschluss daran folgten offene Fragen. Anschließend wurden die Kinder gefragt, ob sie die auf offene Fragen genannten Aktivitäten tatsächlich durchgeführt hatten oder ob sie sie sich vorgestellt hatten. Aktivitäten, die nicht auf offene Fragen benannt wurden, wurden in Frage gestellt, wobei das Interview auch spezifische suggestive Fragen und negativ formulierte beeinflussende Fragen beinhaltete. Insgesamt zeigten sich in dieser Studie kaum Unterschiede zwischen den Gruppen. Bei offenen Fragenformaten zeigten alle Kinder gleichwertige Leistungen, allerdings erzielten Kinder mit durchschnittlicher Intelligenz bei spezifischen Fragen bessere Leistungen. Nach GORDON et al. ist die Ungenauigkeit, mit der geistig behinderte Kinder spezifischen Fragenformaten begegnen auf ein kognitives Defizit zurückzuführen. Im Allgemeinen nahm die Erinnerungsfähigkeit nach sechs Wochen ab, wobei die Forscher im Hinblick auf zukünftige Forschung längere Befragungsverzögerungsintervalle vorschlugen, da dies eher der forensischen Praxis und Realität entspreche. Allen Kindern fiel die korrekte Erinnerung an selbst ausge-

führte Handlungen leichter als an solche, die nur vorgestellt waren, wobei auch fraglich ist, inwiefern die Kinder der Instruktion folgten. Bei den vorgestellten Handlungen benötigten die Kinder Hinweise in Form von spezifischen, geschlossen gestellten Fragen. Zudem erinnern Kinder mit geistiger Behinderung bei offenen Fragen zu einem weniger Inhalt als bei spezifischen Fragen - diese Informationen allerdings genauer - zu anderen wurden Fragen, die Aktivitäten, die nicht auf einer realen Wahrnehmungsgrundlage basierten, betrafen kaum erinnert. Daher schlagen die Forscher vor, dass Kinder mit Entwicklungsverzögerung mittels offener Fragen befragt werden sollten und kindliche Antworten auf spezifische Fragen bezüglich ihres Entwicklungsalters, und nicht ihres Lebensalters, bewertet werden sollten. Des Weiteren blieb die Erinnerung an erlebte Ereignisse über das Verzögerungsintervall bei beiden Gruppen stabiler.

Dies deckt sich mit den Ergebnissen von GOODMAN et al., 1990, wonach bedeutende (erlebte) Ereignisse besser erinnert werden als weniger bedeutende Ereignisse (vorgestellte). Der am häufigsten auftretende Fehlertyp bestand in der fälschlichen Behauptung der befragten Kinder, sie hätten eine Handlung wirklich ausgeführt obwohl sie sich diese nur vorgestellt hatten – dieses Ergebnis war unabhängig vom Intelligenzniveau. Auf die Frage „Haben wir das wirklich gemacht oder hast du dir das nur vorgestellt?“ korrigierten die Kinder ihre anfängliche Antwort. Im zweiten Befragungszeitpunkt korrigierten die Kinder auf diese Nachfragen weniger oft ihre Antwort. Insgesamt ergaben sich jedoch kaum Gruppenunterschiede hinsichtlich der Suggestibilität von geistig behinderten Kindern und der normal begabten Kontrollgruppe. Beide Gruppen von Kindern zeigten sich weniger suggestibel bei Fragen, die sich auf Ereignisse bezogen, die tatsächlich passierten. Zusammenfassend ergab die Studie, dass die Gedächtnisleistungen aller Kinder relativ gut waren. Es wurden nur wenige Unterschiede zwischen den Kindern mit einer Intelligenzminderung und den Kindern im selben Entwicklungsalter gefunden. Kinder mit einer Intelligenzminderung waren nur bei den spezifischen Fragen signifikant schlechter. Die Ergebnisse dieser Studie implizieren für die forensische Arbeit, dass Kinder mit einer Intelligenzminderung in der Lage sind, zuverlässige Aussagen bei polizeilichen oder gerichtlichen Befragungen zu liefern. Ihre Leistungen sind vergleichbar mit denen von Kindern im selben Entwicklungsalter.

LORSBACH (2000) verweist auf eine weitere Untersuchung bezüglich der Fähigkeit zur internalen Quellendiskrimination bei Kindern mit Entwicklungsstörungen. Jens, Gordon & Shaddock (1990) verglichen dabei geistig behinderte Kinder im Alter von 10 Jahren mit normal begabten 6-Jährigen hinsichtlich ihrer Erinnerungs- und Diskriminationsleistung von selbst ausgeführten und nur vorgestellten Aktivitäten. Die Hälfte der Kinder wurde auch ein zweites Mal nach acht Wochen dazu befragt. Im Ergebnis fanden sich keine Gruppenunterschiede hinsichtlich Defizite bei Quellenidentifikationen, wobei beide Gruppen ähnliche

Antwortmuster zeigten. Dabei erweis sich die Quellenzuordnung für selbst ausgeführte Aktivitäten im Allgemeinen als genauer als für zuvor nur vorgestellte.

Dagegen fand LORSBACH in seinen Untersuchungen übereinstimmende Befunde, derart, dass Kinder mit Lernbehinderungen Schwierigkeiten haben, die Quellen ihrer Erinnerungen zu unterscheiden (LORSBACH & EWING, 1995; LORSBACH et al., 1998), da sie möglicherweise Probleme in der reflektiven Verarbeitung und bei verbalen Aufgaben haben (LORSBACH, 2000).

Benetto et al. (1996, zit. nach LORSBACH, 2000) untersuchten 'source monitoring' und Gedächtnisleistung bei autistischen Jugendlichen anhand verschiedener Erinnerungs- und Wiedererkennungsaufgaben. Dabei begingen die autistischen Jugendlichen bei der Erinnerungsaufgabe mehr Fehler in Form von Intrusionen als die nicht-autistische Kontrollgruppe. Ebenso fanden sich bei den recognition-Tests höhere false-alarms-Raten für semantisch ähnliche Items aus den Listen. Die Autoren folgern daraus, dass die autistischen Jugendlichen ein Defizit hinsichtlich der Quellenzuordnung haben und nicht oder nur schwer in der Lage sind, den Kontext der aktuellen Aufgabe zu nutzen, um ungeeignete Antwortalternativen auszuschließen.

Hinsichtlich der Fähigkeiten der Realitätsüberwachung kamen Farrant et al. in ihrer Studie (1998, zit. nach LORSBACH, 2000) beim Vergleich autistischer, geistig behinderter und normal begabter Kinder zu dem Ergebnis, dass geistig behinderte Kinder in Wiedererkennungsleistungen schlechter abschnitten als autistische und normalbegabte Kinder, die in etwa vergleichbare Leistungen zeigten. Keinerlei Gruppenunterschiede konnten jedoch hinsichtlich der Realitätsüberwachung gefunden werden.

LIBBY (2003) konnte zeigen, dass die Bildperspektive - erste Person (selbst) gegenüber dritter Person (Beobachter) - die Beurteilung der Quellenidentifikation beeinflusst (imaginäre Inflation). Diese tritt dann auf, wenn bildhafte Erfahrungen mit Ereignissen fälschlicherweise als reale Erfahrung missverstanden werden. Imaginäre Inflation kommt nicht nur durch die bildhafte Repräsentation selbst zustande, sondern auch durch die Interaktion von Merkmalen dieser Repräsentation und den bei der Abfrage angewandten Quellenidentifikationskriterien. Unterschiede im Ausmaß der Veränderung von Gedächtnisinhalten durch Vorstellung können somit nicht einfach aufgrund von Unterschieden der visuellen Perspektive erklärt werden, die eine Versuchsperson gebrauchte, um sich eine bildhafte Repräsentation der Vergangenheit zu konstruieren. Die Erwartungen an die visuelle Perspektive von Erinnerungen an aktuelle Ereignisse sind dabei von Bedeutung und verändern sich in Abhängigkeit von gegenwärtigen Einflüssen (vgl. LIBBY, 2003).

Obwohl die Zahl der Studien zur Quellenidentifikation begrenzt ist, kann nach dem gegenwärtigen Forschungsstand davon ausgegangen werden, dass lernbehinderte Kinder Schwie-

rigkeiten haben, bewusst Gedächtnisquellen zu erinnern, die verbale Informationen beinhalten. Daraus kann jedoch nicht einfach auf ein grundsätzliches Gedächtnisdefizit geschlossen werden. Vielmehr deuten die Befunde eher auf ein grundlegendes Problem im Gebrauch der von JOHNSON et al. eingeführten Begriffe und Phänomene der heuristischen und systematischen Entscheidungsprozesse hin (vgl. LORSBACH, 2000).

Im Zusammenhang mit all den genannten Untersuchungen ist es allerdings wichtig darauf hinzuweisen, dass es sich zum einen hierbei um experimentelle Paradigmen handelt, zum anderen nicht besonders viele Untersuchungen dazu existieren. Allgemeingültige und schlussfolgernde Befunde gestalten sich demnach zum gegenwärtigen Zeitpunkt der Forschung noch schwierig. Inwiefern bzw. ob sich Leistungsniveaus in forensisch relevanten Situationen aufgrund persönlicher Involviertheit, emotionalem Bezug oder anderen sozialen und motivationalen Faktoren verbessern oder verschlechtern, kann auch als Funktion der Kompetenz des Befragenden verstanden werden.

Entscheidende weitere Voraussetzung einer guten Gedächtnisleistung bzw. für das spätere Erinnern ist die Art und Weise sowie die Qualität der Informationsverarbeitung. Diese setzt sich aus bestimmten Basis-Kompetenzen, Strategien, dem metakognitiven Wissen über spezifische Strategien, motivationalen Bedingungen und anderem nicht-strategischem Wissen zusammen. Da jedoch bei lernbehinderten Menschen davon auszugehen ist, dass einige oder alle der genannten Bereiche/Aspekte gestört oder eingeschränkt sind, sollen diese im Sinne einer Art 'Negativdarstellung' im Folgenden kurz dargestellt werden (vgl. PRESSLEY, 1991).

Hinsichtlich einer intakten und erfolgreichen Informationsverarbeitung gelten als *Basis-Kompetenzen*:

- (1) intakte sprachliche Fähigkeiten und Fertigkeiten mit umfassendem rezeptivem und produktivem Wortschatz
- (2) große Mengen an Informationen und umfassendes Wissen, das im Langzeitgedächtnis gespeichert ist und wovon vieles bei Bedarf zugänglich ist bzw. abgerufen werden kann und den Gebrauch von Strategien ermöglicht, falls entsprechende detaillierte mentale Repräsentationen verfügbar sind
- (3) symbolhafte Veränderungen als Teil von ausführenden Strategien (Wiederholen von Informationen, Aufbau von Bildern, Reorganisation von Input) zeigen sich vor allem im Kurzzeitgedächtnis, weshalb dieses bei Ausführung vieler Strategien, welche bewusste Reflektionen und Veränderungen der inhaltlichen Bedeutung erfordern, notwendig ist
- (4) Fähigkeit zur angemessenen und selektiven Aufmerksamkeit und Konzentration bezüglich Informationen und deren Relevanz

Im Rahmen des *Gebrauchs von Strategien* sind für eine gute Informationsverarbeitung eine Vielzahl derer in unterschiedlichen Bereichen von Bedeutung:

- *Gedächtnis:*
Strategien der Wiederholung, Organisation sowie elaborative Strategien
- *Begriffsvermögen:*
Zusammenfassen von Textinformationen, Bildung repräsentativer Bilder, Aktivierung relevanten Vorwissens
- *allgemein:*
Überprüfung und Überwachung der Ausführung, Versuch des Findens von Ähnlichkeiten von unbekannten und bekannten Aufgaben, Suche nach Hinweisreizen aus der Umwelt, die bedeutsame Informationen signalisieren

Ebenso spielen *motivationale Überzeugungen* eine wichtige Rolle. Zwar kann bei allen Kindern anfänglich von einer intrinsischen Motivation zu lernen und sich zu entwickeln ausgegangen werden: jedoch verstehen Menschen, die über eine erfolgreiche Informationsverarbeitung verfügen, besser, dass es im Rahmen einer normalen kognitiven Entwicklung gilt, verschiedene Strategien zu beherrschen.

Ergänzend und zusammenfassend kann bezüglich lernbehinderter Menschen als schlechtere Informationsverarbeiter neben den bereits oben ausgeführten Defizitcharakteristika festgestellt werden, dass diese oftmals mangelnde Aufmerksamkeit und Impulsivität zeigen. Des Weiteren können Angst hinsichtlich schulischer oder beruflicher Angelegenheiten sowie im Falle einer gegebenenfalls vorhandenen Wissensbasis Probleme des Zugangs und der Anwendung auftreten (vgl. PRESSLEY, 1991).

Auf dem Hintergrund von erhöhten Verbrechensraten gegenüber dieser speziellen Personengruppe und auch Kindern, scheint es angebracht, Kinder und Jugendliche aus unterschiedlichen familiären und intellektuellen Verhältnissen und Bereichen in die Forschung der Quellenüberwachung mit einzubeziehen, um genaue Vorstellungen darüber und angemessene Interviewtechniken entwickeln zu können.

1.4.1.3 Gedächtnisleistungen - Suggestibilität und Beeinflussbarkeit bei Intelligenzminderung

Hinsichtlich des Zusammenhangs zwischen *Intelligenz* und Suggestibilität scheint die Befundlage inkonsistent, da sich hier zum Teil negative Korrelationen ergaben (Gudjonsson, 1988, 1990; Gudjonsson & Sigurdson, 1996; zit. nach GREUEL, 2001), zum Teil jedoch

keine Zusammenhänge gefunden wurden (GUDJONSSON, 1992; GODJONSSON & CLARE, 1995). Bei intellektuell minderbegabten Personen waren diese allerdings deutlich vorhanden (GUDJONSSON, 1992). Wenn sich also insoweit feststellen lässt, dass intellektuelle Minderbegabung mit erhöhter Suggestionsanfälligkeit korrespondiert, bliebe jedoch immer noch zu klären, ob es sich hierbei einen originär kognitiv begründeten Effekt handelt oder ob dabei nicht vielmehr soziale Bedingungen (verringertes Selbstbewusstsein, erhöhte compliance) wirksam werden (vgl. GREUEL, 2001). So ist verschiedentlich darauf hingewiesen worden, dass minderbegabte Personen spezifische Sozialisationserfahrungen machen, indem sie lernen, dass ihre kognitiven Leistungen geringer sind als die anderer und eine Anpassung an und Übernahme der Vorgaben Dritter (oft autoritär oder zumindest kompetent(er) wirkender Personen) daher vor Bloßstellung schützen kann. Insofern begünstigen derartige Lernerfahrungen nicht nur eine erhöhte Anfälligkeit für interrogative Suggestionen, sondern erhöhen auch das individuelle Viktimisierungsrisiko (Dettenborn et al., 1989; Sgroi, 1989; Horton & Kochurka, 1995; zit. nach GREUEL, 2001).

In diesem Zusammenhang und der Analyse von Persönlichkeitsmerkmalen ist hier wiederum gerade bei Kindern und Jugendlichen mit intellektuellen Beeinträchtigungen zum einen der *erhöhten Ängstlichkeit* im Sinne einer *situationsgebundenen 'state anxiety'* von großer Bedeutung (*situative Suggestivität*) für die (suggestionsbedingte) Veränderung von Aussageinhalten. Dabei korreliert nach GREUEL (2001) die Höhe der situativ empfundenen Ängstlichkeit insbesondere mit der Vulnerabilität für interpersonalen Druck (shift). Schüchterne Kinder empfinden demnach generell bei konkreten Fragen eher Befragungsdruck, der dazu führen kann, emotionsvermeidende, palliative Coping-Strategien zur Bewältigung der Situationsanforderungen anzuwenden.

Ebenso gehen ein geringes Selbstwertgefühl – insbesondere hinsichtlich der Selbstattribution von Kompetenz und Kontrolle – mit einer Erhöhung der Befragungs-Suggestibilität einher, vor allem was die Empfänglichkeit für negatives Feedback betrifft. Dies gilt insbesondere für Personen, die insgesamt wenig mit Befragungssituationen und entsprechenden Umständen vertraut sind. Analoge Zusammenhänge wurden bezüglich externaler Kontrollüberzeugungen und Feldabhängigkeit nachgewiesen (Blagrove et al., 1994; Gudjonsson & Clark, 1986; Singh & Gudjonsson, 1992; zit. nach GREUEL, 2001). Somit kann angenommen werden, dass die genannten Persönlichkeitsmerkmale gerade in solchen Befragungssituationen eine Mediatorwirkung haben können, in denen eine Distanz zwischen Befrager und Befragtem im Sinne einer asymmetrischen Beziehungs- und Kommunikationsform gegeben ist.

So lag der Focus einer Studie von CLARE & GUDJONSSON (1993) auf der Frage, ob Menschen mit einer Lernbehinderung in der Lage sind, eine zuverlässige Aussage zu machen, wobei erhoben wurde, ob diese Personengruppe suggestibler ist als Personen mit durchschnittlichen Fähigkeiten, ob sie mehr konfabuliert und ob sich eine Tendenz zeigt, Fragen mit „Ja“ zu beantworten unabhängig ihres Inhalts. Zum einen umfasste die Suggestibilität, irreführenden Fragen nachzugeben („yield“) und zum anderen die Veränderung einer ursprünglichen Antwort als Folge von negativer Rückmeldung („shift“). Konfabulationen wurden definiert als verzerrte Quellenidentifikationsleistung, d.h. eine Erinnerung, bei der Erinnerungslücken mit Einfällen gefüllt werden, die vom Befragten selbst für Erinnerungen gehalten werden. Abschließend wurde die Tendenz erhoben, Fragen unabhängig ihres Inhalts bejahend zu beantworten. Insgesamt ergab sich, dass die Versuchspersonen mit einer Intelligenzminderung in allen untersuchten Bereichen (Suggestibilität, Konfabulationen, Zustimmungstendenz) sich diesbezüglich als anfälliger erwiesen als die normalbegabten Personen aus der Kontrollgruppe und zudem schlechtere Erinnerungsleistungen zeigten. Im Detail ließen sich die Versuchspersonen mit einer Lernbehinderung durch „leitende Fragen“ mehr beeinflussen als die Kontrollgruppe. Entgegen den Erwartungen unterschieden sich die Gruppen nicht in ihrer Leistung bei einem „negativ gegebenen Feedback“. Zudem fanden sich Konfabulationen und Zustimmungstendenz bei Menschen mit einer Lernbehinderung signifikant häufiger als bei Menschen ohne eine intellektuelle Beeinträchtigung. Die Tendenz, Fragen bejahend zu beantworten, korrelierte negativ mit der allgemeinen intellektuellen Fähigkeit. Zudem stellte sich heraus, dass schlechtere Erinnerungsleistungen mit einer höheren Anzahl an Konfabulationen korrelierten. Die Autoren schlussfolgern daraus, wenn Menschen mit Lernbehinderung in höherem Ausmaß konfabulieren als andere mit durchschnittlichen Fähigkeiten, dann machen sie selbst unter optimalen Befragungsbedingungen weniger zuverlässige Aussagen (vgl. CLARE & GUDJONSSON, 1993).

Demgegenüber entgegnet BEAIL (2002) kritisch, dass einige Forschungsergebnisse, insbesondere diejenigen, die auf der standardisierten Gudjonsson Suggestibility Scale (GSS) beruhen, nur begrenzt auf den juristischen Kontext anwendbar sind, da Gedächtnisexperimente oft nur das semantische Gedächtnis prüfen, wohingegen bei gerichtlichen oder polizeilichen Befragungen das episodische Gedächtnis angesprochen wird und dass sich die Erfahrung, eine Geschichte vorgelesen zu bekommen und diese erinnern zu müssen, völlig vom persönlichen Erleben eines Ereignisses und der Wiedergabe desselben unterscheidet (vgl. BEAIL, 2002). Die Autoren der angeführten Studien weisen auch immer wieder darauf hin, dass es in einem erhöhten Maß auf die Fragetechnik ankommt, ob ein Zeuge vulnerabel ist oder nicht.

So war es das Anliegen und Ziel der Untersuchung von PERLMAN et al. (1994) diejenigen Bedingungen zu erheben, unter welchen Menschen mit kognitiven Defiziten in der Lage sind,

eine zuverlässige und genaue Aussage zu machen. Insgesamt ist nach PERLMAN und Mitarbeitern davon auszugehen, dass Menschen mit Entwicklungsverzögerungen unter bestimmten Bedingungen kompetenter, weniger kompetent oder genauso kompetent in ihrer Fähigkeit sein können, Ereignisse zu berichten. Die genannte Arbeitsgruppe zeigte hierzu einen kurzen Film und verglich anschließend das Antwortverhalten von 30 kognitiv beeinträchtigten Menschen und 30 normalbegabten Personen auf unterschiedliche Fragenformate mit zum Teil richtig bzw. falsch beeinflussenden Inhalten. Insgesamt zeigte sich, dass die Kontrollgruppe bei den unstrukturierten Fragenformaten - freier Bericht und allgemeine Fragen - etwa doppelt so viele korrekte Items erinnerte wie die Befragten mit Entwicklungsverzögerung. Allerdings unterschieden sich beide Gruppen nicht in Bezug auf die Menge der falsch wiedergegebenen Information. So konnte festgestellt werden, dass die Befragten mit Entwicklungsverzögerung zwar nicht so detailreich erinnerten wie die Kontrollgruppe, allerdings war bei beiden Gruppen etwa 80% der wiedergegebenen Information korrekt. Im Gegensatz dazu erinnern in der Untersuchung von GUDJONSSON & CLARE (1993) die intellektuell beeinträchtigten Versuchspersonen durchschnittlich nur 25 % korrekte Informationen beim unmittelbaren freien Bericht. Auch MILNE & BULL (2001) bestätigen diese Tendenz, dass intellektuell beeinträchtigte Menschen zwar nicht so detailliert, aber dennoch korrekt erinnern.

Ein weiteres Ergebnis von PERLMAN et al. (1994) bestand in dem Befund, dass bei den strukturierten Fragenformaten - spezifische suggestive Fragen und beeinflussende Ja/Nein-Fragen - Befragte mit Entwicklungsverzögerung weniger vollständige und richtige Informationen wiedergaben als die Kontrollgruppe. Dies deckt sich mit dem Befund von DENT (1986), wobei die Genauigkeit der Erinnerung an ein erlebtes Ereignis von geistig behinderten Kindern untersucht wurde. Demnach seien Bedenken an der Angemessenheit dieser Fragenformate für Menschen mit Entwicklungsverzögerungen angebracht. Weiterhin erinnerten beide Versuchsgruppen zum einen Details und Informationen, die sich auf das zentrale Kerngeschehen bezogen, besser, zum andern konnte bei den Versuchspersonen mit Entwicklungsverzögerung ein verstärktes Bedürfnis nach Konformität mit den antizipierten Wünschen einer Autoritätsfigur beim Antwortverhalten auf fehlerleitende, strukturierte Fragenformate festgestellt werden (vgl. PERLMAN et al., 1994). Die Autoren empfehlen, dass es bei der Befragung im juristischen Kontext durchaus erforderlich ist, Zeugen mit Entwicklungsverzögerung über ihr Recht aufzuklären, die Behauptungen von Anwälten zurückzuweisen, um somit der Gefahr einer erhöhten 'compliance' gegenüber den wahrgenommenen Autoritätspersonen entgegenzuwirken. Zudem müssten Richter über Bedingungen aufgeklärt werden, unter welchen Bedingungen Zeugen mit Entwicklungsverzögerung anfällig sein können für Suggestion. Allgemein wird ein Fragenformat des freien Berichts in Kombination

mit nicht beeinflussenden Ja/Nein-Fragen vorgeschlagen, um möglichst viel genaue Information über diese Personengruppe zu erhalten.

Auch HENRY & GUDJONSSON (1999) verglichen intellektuell beeinträchtigte Kinder und Kinder gleichen Entwicklungsalters sowie Kinder gleichen Lebensalters hinsichtlich ihrer Fähigkeit, ein erlebtes Ereignis einen Tag später wiederzugeben in Abhängigkeit von unterschiedlichen Fragetechniken und dem Suggestibilitätsniveau. Bisherige Ergebnisse aus entsprechenden Studien deuten darauf hin, dass Kinder mit einer Intelligenzminderung Informationen genauso gut enkodieren, speichern und abrufen können wie Kinder des gleichen Entwicklungsalters, sofern es sich nicht um künstlich erzeugte Gedächtnisexperimente handelt. Ein automatisiertes Erinnern scheint leichter für diese Personengruppe zu sein, da es keine Gedächtnisstrategien wie z.B. inneres Vorsagen benötigt (vgl. HENRY & GUDJONSSON, 1999). Im Rahmen der genannten Untersuchung ist zu beachten, dass die Gruppe der intellektuell beeinträchtigten Kinder sehr unterschiedliche Intelligenzniveaus aufwies; so weist etwa die Hälfte der Kinder eine geistige Behinderung auf, die restlichen Kinder sind lernbehindert. Die Kinder wurden mittels eines festgelegten Fragenkatalogs zu einer Tags zuvor präsentierten Szene befragt, welcher eingangs freien Bericht beinhaltete, dann allgemeine Fragen, offene spezifische Fragen, wobei die Hälfte irreführend war, die andere Hälfte nicht. Ferner folgten geschlossene Ja/Nein-Fragen, wobei wiederum die Hälfte irreführend formuliert war, die andere Hälfte richtige Details vorgab. Zudem wurde die bereits erwähnt und auch kritisierte GSS (Gudjonsson Suggestibility Scale) angewendet. In Übereinstimmung mit bisher ähnlichen Befunden zeigt sich auch hier, dass sich die Leistung von intellektuell beeinträchtigten Kindern und normal begabten Kindern kaum unterschied. So fanden sich keine signifikanten Unterschiede bezüglich offener Fragen oder richtig lenkenden Ja/Nein-Fragen. Allerdings war der Umfang des freien Berichts bei intellektuell beeinträchtigten Kindern geringer und auf allgemeine Fragen erfolgten keine weiteren Informationen. Der Hauptunterschied zwischen den intellektuell beeinträchtigten Kindern und der gleichaltrigen Kontrollgruppe lag in der Reaktion auf irreführende geschlossene Ja/Nein-Fragen, da sich die Kinder der Experimentalgruppe hier suggestibler zeigten und somit der Leistung der Kontrollgruppe gleichen Entwicklungsalters entsprachen. Neben kognitiven Faktoren sind nach Henry & Gudjonsson in erster Linie soziale Faktoren für die erhöhte Suggestibilität intellektuell beeinträchtigter Kinder verantwortlich. Demnach führen geschlossene fehlleitende Fragen zu starkem Druck bei diesen Kindern, mit dem Interviewer übereinzustimmen bzw. ihn zufrieden zu stellen. Vermindertes Selbstbewusstsein sowie Vertrauen in die eigenen Fähigkeiten und ein Widerstreben, einer Autoritätsperson zu widersprechen, werden für diesen Effekt verantwortlich gemacht. Demnach stellt das Entwicklungsalter eines Zeugen einen guten Prädiktor für das Suggestibilitätsniveau dar. Zudem geben die Autoren

zu bedenken, dass sich ein fehlerleitender geschlossener Fragestil bei intellektuell beeinträchtigten Zeugen problematisch erweisen könnte (vgl. HENRY & GUDJONSSON, 1999).

Einige Jahre später lag der Focus von HENRY & GUDJONSSON in einer Untersuchung von 2003 auf dem Augenzeugengedächtnis wiederum von Kindern im Alter von 11 bis 12 Jahren mit einer leicht- bis mittelgradigen Intelligenzminderung (IQ's 45-79). Als Kontrollgruppen fungierten Kinder gleichen Entwicklungsalters und auch vergleichbaren Lebensalters. Aufgrund der Relevanz für den forensischen Kontext wurden darüber hinaus die Auswirkungen wiederholter Befragungen erhoben, da es durchaus üblich ist, dass Zeugen mehrmals über dasselbe Ereignis befragt werden. Zudem sollte untersucht werden, inwiefern standardisierte Maße geeignet sind, die Zeugenkompetenz zu bestimmen. Hinsichtlich Reizmaterial und Befragungsmodalitäten analog zur oben dargestellten Studie von 1999 wurden die Kinder einen Tag sowie zwei Wochen nach Stimuluspräsentation befragt. Die Untersuchung ergab beim Vergleich von lernbehinderten Kindern und Kindern gleichen Lebensalters, dass intellektuell beeinträchtigte Kinder ähnliche Leistungen beim freien Bericht zeigten wie normal begabte Kinder und sich zudem nicht als anfälliger für Suggestionen erwiesen, was den Ergebnissen aus dem Jahre 1999 widerspricht. Jedoch waren lernbehinderte Kinder beim freien Bericht und auf allgemeine Fragen weniger in der Lage, korrekte Informationen zu erinnern als Kinder der gleichaltrigen Kontrollgruppe. Zudem änderten die Kinder mit Lernbehinderung ihre Antworten auf spezifische Fragen in der zweiten Befragungsbedingung signifikant häufiger. Henry & Gudjonsson schließen daraus, dass sich intellektuell beeinträchtigte Kinder durch die im forensischen Kontext üblichen wiederholten Befragungen stärker beeinträchtigt und verunsichert fühlen. Trotzdem kann daraus nicht pauschal von einer schlechteren Gedächtnisleitung oder Erinnerungsfähigkeit von lernbehinderten Kindern ausgegangen bzw. diese erwartet werden. Jedoch weisen die Autoren beim Vergleich von geistig behinderten Kindern und Kindern gleichen Lebensalters durchaus auf eine erhöhte Vulnerabilität der Personengruppe mit intellektueller Beeinträchtigung hin. Kinder mit geistiger Behinderung erzielten in fast jedem Aspekt geringere Leistungen und weniger Informationsdichte als gleichaltrige Kinder bezüglich des freien Berichts, bei allgemeinen und auch offenen nicht-suggestiven Fragen sowie geschlossenen irreführenden Fragen. Zudem änderten die geistig behinderten Kinder ihre Antworten in der zweiten Befragungssituation häufiger. Für diese Ergebnisse werden - wie schon mehrmals angedeutet - auch hier vor allem soziale Faktoren verantwortlich gemacht. Der Vergleich mit Kindern gleichen Entwicklungsalters erlaubt nun generell die Untersuchung, ob die Leistung von intellektuell beeinträchtigten Kindern mit ihrem Entwicklungsalter Schritt hält oder dahinter fällt. Zum einen stellte sich der positive Effekt heraus, dass Kinder mit intellektueller Beeinträchtigung im freien Bericht mehr exakte Informationen erinnerten als die Kontrollgruppe gleichen Entwicklungsalters. Zum anderen ergab sich jedoch auch der negativ zu bewertende Effekt, dass

Kinder mit intellektueller Beeinträchtigung ihre Antworten bei der zweiten Befragung häufiger änderten als Kinder ähnlichen Entwicklungsstands. Die Autoren vermuten, dass sich intellektuell beeinträchtigte Kinder beeinflussbarer zeigten durch die antizipierte Unangemessenheit ihrer ursprünglichen Antworten, bedingt durch eine erneute, gleichartige Befragungssituation. Die vermehrten Änderungen im Aussageverhalten ('eyewitness shift') stellen nach Ansicht von Henry & Gudjonsson scheinbar eine besondere Schwäche von Kindern mit einer Intelligenzminderung da. Demnach sind Mehrfachbefragungen bei Zeugen zumindest kritisch bzw. problematisch anzusehen. Gudjonsson und Clark, führen zwei Gründe an, warum Menschen mit einer Intelligenzminderung in einem erhöhten Maße suggestibel sind: Zum einen durch eine schwächere Gedächtniskapazität ('kognitiver Faktor') und zum anderen durch eine geringere Fähigkeit sich vor Unsicherheiten und Erwartungen bei Befragungen zu schützen ('sozialer Faktor') (zit. nach HENRY & GUDJONSSON, 2003). Interessanterweise konnten die negativen Konsequenzen einer wiederholten Befragung unabhängig von einer intellektuellen Beeinträchtigung beobachtet werden. Zusammenfassend lässt sich sagen, dass Kinder mit einer leichten Intelligenzminderung im freien Bericht und bzgl. der Suggestibilität keine Unterschiede zu normal begabten Kindern im selben Lebensalter aufweisen. Nach HENRY & GUDJONSSON (2003) bleibt jedoch offen, inwieweit die Ergebnisse auch auf reale, strafrechtlich relevante, beobachtete Handlungen übertragbar sind.

Zusammenfassend zeigen die hier angeführten Studien zur Zeugenkompetenz bei Menschen mit einer Intelligenzminderung, dass von einer generellen Zeugenuntüchtigkeit nicht ausgegangen werden kann. Defizite sind nur dann offensichtlich geworden, wenn gleichaltrige Personen den Maßstab bilden. In den zitierten Studien, bei denen Menschen im selben Entwicklungsalter die Vergleichstichprobe darstellen, konnten keine manifesten Unterschiede gefunden werden. Es hat sich demnach als hilfreich erwiesen, die Zeugenleistungen von Kindern bzw. Erwachsenen mit einer Intelligenzminderung mit der von Menschen die im gleichen Entwicklungsalter sind zu vergleichen. Da normal begabte Kinder über 4 Jahre in der Regel als aussagetüchtig angesehen werden, kann davon ausgegangen werden, dass auch Menschen mit einer Intelligenzminderung aussagetüchtig sind, sofern das Entwicklungsalter eines 4-Jährigen gegeben ist. Dabei müssen auch bei diesem Personenkreis die allgemeinen Vorraussetzungen wie z.B. eine zuverlässige Wahrnehmung des Ereignisses, ein ausreichendes Sprachverständnis etc. gegeben sein.

In Ergänzung zu den oben ausgeführten Simulationsstudien zur Zeugenkompetenz intellektuell beeinträchtigter Menschen sei trotz der Unterschiede in der Vernehmungs- und Befragungspraxis in deutschen und angelsächsischen Rechtssystemen noch auf eine aktuelle Feldstudie von KEBBELL et al. (2004) verwiesen, da auch sie Implikationen zur Gestaltung

optimaler Befragungsbedingungen geben kann. In der Untersuchung wurden Transskripte von Gerichtsverhandlungen von Sexualdelikten gegen Zeugen mit und ohne intellektuelle Beeinträchtigung verglichen, die Befragungsstrategien analysiert und deren Auswirkungen eruiert. KEBBELL et al. (2004) fanden heraus, dass Anwälte ihr Frageverhalten weder positiv noch negativ bei intellektuell beeinträchtigten Zeugen veränderten. Bedenkt man, dass das juristische Ausdrucksverhalten oftmals von Fachtermini, komplexem Satzbau und doppelten Verneinungen geprägt ist und intellektuell beeinträchtigte Menschen damit Schwierigkeiten haben können, so ist es durchaus möglich, dass zahlreiche Befragungen suboptimal verlaufen. Außerdem wurden Fragen bei dieser Personengruppe öfter wiederholt, was laut den Autoren entweder an einer suggestiblen Strategie liegen kann oder an der Tatsache, dass die Zeugen nicht verstanden wurden. Beim Kreuzverhör wurden insgesamt signifikant mehr Ja/Nein-Fragen, negativ formulierte Fragen, multiple Fragen (Fragen, die mehrere Aspekte umfassen) und lenkende Fragen gestellt als bei der direkten Befragung, wobei sich derartige Fragenformate als problematisch erwiesen, da intellektuell beeinträchtigte Zeugen hier weniger Informationen gaben als normal begabte Befragte. Ferner war festzustellen, dass sich Zeugen mit intellektueller Beeinträchtigung anfälliger zeigten, einer lenkenden Frage nachzugeben und weniger fähig, einer derartigen Frage zu widersprechen. Die Autoren folgern, dass derartig praktizierte Befragungen durchaus zu ungenauen Zeugenaussagen führen können und betonen das Erfordernis, dass Anwälte und Richter über die Besonderheiten intellektuell beeinträchtigter Zeugen und angemessene Befragungsstrategien aufgeklärt werden. Generell sollten alle notwendigen, empirisch-theoretischen und praxisbezogenen Schritte unternommen werden, um den Zugang lernbehinderter Menschen zu Gerichten zu verbessern und die stereotypisierte Ansicht und Meinung über diese Menschen abzuschaffen (vgl. CLARE & MURPHY, 2001).

Dies impliziert und verlangt somit weitere ökologisch valide Forschungsarbeiten auf dem Gebiet der Psychologie der Zeugenaussage, insbesondere bei Menschen mit intellektuellen Beeinträchtigungen, um im Rahmen von polizeilichen und/oder gerichtlichen Vernehmungen, aber natürlich auch im Bereich der Begutachtung durch Sachverständige mehr Klarheit im Umgang mit und bessere Einschätzung der jeweiligen Kompetenzen dieser speziellen Personengruppe zu gewinnen. Ziel der Forschungsarbeiten sollten Befunde sein, die die Grundlage darstellen, um Befragungsbedingungen konkret so zu gestalten, um möglichst optimale Abruf-, Erinnerungs- und Gedächtnisleistungen von Menschen mit kognitiven Defiziten zu gewährleisten bzw. zu produzieren.

1.4.2 Interviewtechniken und Implikationen für eine optimale Befragungssituation

Aus der entsprechenden Forschung bei normal begabten Personen (Kinder und Erwachsene) lassen sich einige, jedoch nicht alle Befunde auf den Personenkreis der Lernbehinderten übertragen bzw. anwenden. Die folgenden Ausführungen nehmen teilweise Bezug darauf, sollen aber speziell für die genannte Personengruppe gelten.

Zu Beginn einer Interviewsituation sollte zuerst einmal genügend Zeit aufgewendet werden, um *Rapport aufzubauen*, da sich intellektuell beeinträchtigte Menschen in ihren Kompetenzen, Bedürfnissen und Grenzen stark von normal begabten Menschen unterscheiden. Dies hat zum einen den Zweck, den Interviewten kennen zulernen und negativen Erwartungshaltungen und Ängsten entgegenzuwirken, die unter Umständen aufgrund früherer negativer Erfahrungen bestehen (vgl. MILNE & BULL, 2001; COOKE & DAVIES, 2001). Dies kann beispielsweise dadurch erreicht werden, dass die Gründe für das Interview erläutert werden und das Verständnis des Befragten für die Situation erhoben wird. Zum anderen können in der Phase des Rapportaufbaus die kognitiven, emotionalen und sozialen Kompetenzen herausgearbeitet und beobachtet werden, um sich in der nachfolgenden Befragungssituation an die besonderen Erfordernisse des Interviewten anpassen zu können (MILNE, 1999). Dieser Rapportaufbau kann bei Menschen mit einer Lernbehinderung 30-40 Minuten in Anspruch nehmen. Stets gilt dabei empathisch auf die Bedürfnisse des Befragten einzugehen und die Fragen auf die kommunikativen Fähigkeiten des Gegenübers abzustimmen (vgl. MILNE & BULL, 2001).

Grundsätzlich sollte in Befragungen darauf geachtet werden, eine *angenehme Atmosphäre* zu schaffen und neben geeigneten, noch genauer zu erläuternden Fragetechniken, den Raum der Befragung frei von Störungen zu halten, da Menschen mit intellektuellen Defiziten oftmals Schwierigkeiten haben, sich an unübliche Situationen anzupassen und sich diese Probleme auf Erinnerungsleistung und Kommunikationsverhalten auswirken können (MILNE, 1999). Bei einer Befragung im juristischen Kontext kann es beispielsweise durchaus sinnvoll und fördernd sein, Zeugen über das Vorgehen bei Gericht genau zu informieren, ihnen den Zweck der Befragungen zu erklären und herauszustellen, dass die Interviewenden selbst nichts über die von intellektuell beeinträchtigten Zeugen zu bekundenden Vorfälle wissen, da diese bei dem fraglichen Ereignis nicht anwesend waren und somit nicht wüssten, was passiert sei. Andernfalls besteht zumindest bei Kindern mit einer Lernbehinderung die Möglichkeit, dass diese die Vermutung haben, der Interviewer sei kundig. (vgl. MILNE, 1999; Bull, 1995). BULL (1995) weist des weiteren darauf hin, dass Kinder mit einer intellektuellen Beeinträchtigung nicht unbedingt verstehen, dass der Befragende *detaillierte Informationen*

benötigt, weshalb es hilfreich und notwendig erscheint, *explizit auf den Bedarf hinzuweisen* und auch zu betonen, dass der Interviewer nicht dabei war und deswegen so genau nach dem Ereignis fragt.

Im Verlauf des der gesamten Befragung ist es nach MILNE & BULL (2001) zudem erforderlich, *langsam zu sprechen* und genügend Zeit für die Befragung einzuplanen. Zudem ist es sinnvoll, das eigene *Kommunikationsverhalten* (z.B. Satzlänge, Ausdrucksverhalten) an die Bedürfnisse des Interviewten *anzupassen* und - soweit möglich - dem Befragten zu vermitteln, dass er stets die Kontrolle über die Situation hat, um sozialer Angst und geringem Selbstvertrauen, das gerade und insbesondere bei Menschen mit Behinderungen aus dem schon an anderer Stelle erläuterten Machtgefälle zwischen Befragtem und Befragendem erwächst, entgegen zu wirken.

MILNE (1999) schreibt, dass Kinder mit einer Lernbehinderung zwar genaue, aber größtenteils nur sehr knappe freie Berichte liefern. Deshalb ist der Interviewer zwangsläufig angehalten, weiterführende Fragen zu formulieren, wobei grundsätzlich und in jedem Fall darauf zu achten ist, dass die angewendete Fragetechnik den Befragten nicht zu sehr in seinem Abruf manipuliert.

Bezüglich der Anwendung von bestimmten Befragungsstrategien weisen einige Autoren darauf hin, dass sich zwei Komplexe von Fragentypen als problematisch erweisen könnten. Zum einen sind Fragen zu vermeiden, die implizit oder explizit die *erwartete Antwort schon vorgeben* und zum anderen empfiehlt es sich, auf Fragen zu verzichten, die den Interviewten *verwirren* könnten. Zu ersterem gehören beispielsweise geschlossene Fragen wie Ja/Nein-Fragen, Entweder/Oder-Fragen, lenkende Fragen, Wahlfragen, Fragen, die eine kurze Antwort erfordern (wenige Wörter) und wiederholte Fragen zu einem Aspekt. (vgl. KEBBEL et al., 2001 ; MILNE & BULL, 2001; PERLMAN et al., 1994). Bei den Ja/Nein-Fragen zeigte sich, dass intellektuell beeinträchtigte Menschen dazu neigen können, in die Frage unabhängig ihres Inhalts einzuwilligen (vgl. GUDJONSSON & CLARE, 1993). Entweder/Oder-Fragen können sich als problembehaftet erweisen, da der Zeuge dadurch fälschlich annehmen könnte, die vorgegebenen Alternativen seien die einzig möglichen (KEBBEL et al., 2001). Vielfach wurde in der Forschung (GORDON et al., 1994; GUDJONSSON & CLARE, 1993) bestätigt, dass sich bei lenkenden, insbesondere irreführenden, Fragen aufgrund oben dargelegter sozialer Faktoren eine erhöhte Suggestibilitätstendenz ergibt, was sich durchaus negativ auf die Zeugeneignung auswirken kann. Auch Wahlfragen sind nicht unbedingt zu verwenden, da im Sinne einer Art 'recency-effect' die Tendenz besteht, in die letzte Option einzuwilligen, möglicherweise auch aufgrund einer Schwierigkeit, alle Antwortmöglichkeiten zu erinnern (PROSSER & BROMLEY, 1998).

Spezifische und geschlossene Fragen erweisen sich somit als grundsätzlich problematisch und sollten daher nur Verwendung finden, wenn alle anderen Fragetechniken bereits Anwendung gefunden haben und ausgeschöpft sind.

Bei Fragen, die eine kurze Antwort erfordern, ist es möglich, dass intellektuell beeinträchtigte Zeugen dazu tendieren, unabhängig von der Richtigkeit eine Antwort zu geben, um nicht eingestehen zu müssen, die Antwort nicht zu wissen (PERLMAN et al., 1994). Auch dieses Frageverhalten erweist sich als kontraproduktiv im Sinne des Generierens einer zuverlässigen Aussage. MILNE & BULL (2001) betonen auch den Aspekt der Wiederholung von Fragen zum gleichen Themenkomplex als kritisch bzw. problematisch, da der Befragte möglicherweise einfach deshalb seine Antworten ändert, da er annimmt, die ursprüngliche Auskunft sei falsch.

Doppeldeutige bzw. nicht eindeutige Fragenformate sollten ebenso vermieden werden, da diese den intellektuell beeinträchtigten Zeugen eventuell verwirren könnten. Nach KEBBELL et al. (2001) erweisen sich Fragen mit Verneinungen oder doppelten Verneinungen, eine komplexe Satzstruktur und unangemessenes Vokabular hier als kritisch und sollten demzufolge vermieden werden. Zudem kann angenommen werden, dass intellektuell beeinträchtigte Zeugen Probleme damit haben, Namen, Zahlen, Daten und Zeiten korrekt zu erinnern und deren Abfrage somit zu Unsicherheit und Verwirrung führen kann (vgl. KEBBELL et al., 2001).

Ebenso wie Fragewiederholungen können Veränderungen der Frageform implizit die Aussage eines Zeugen (negativ) beeinflussen, derart, dass ein Zeuge sich veranlasst fühlt zu glauben, er habe eine „falsch“ Antwort gegeben und diese nun aufgrund der schon diskutierten, verschiedenen Gründe ändern zu müssen oder zu sollen. HUNT & BORGIDA (2001) fanden, dass solche Modifikationen eher ignoriert als korrigiert werden oder diesen zugestimmt wird. Erwachsene stimmten Modifikationen meist nicht zu, während etwa Vorschulkinder am meisten dazu neigten, abgeänderte Falschinformation in nachfolgende Antworten einzubauen. Aufgrund der ausführlich dargestellten sozial, emotional, kommunikativ und interpersonell problematischen Situation (lern-)behinderter Zeugen sowie dem großen, potentiellen Einfluss autoritärer Personen und dem Selbstkonzept dieser Personengruppe, sind in jedem Fall unangemessene Fragestrategien zu vermeiden.

Nachstehende Tabelle fasst auf der Grundlage verschiedener Untersuchungen (KEBBELL et al., 2001; KEBBELL & JOHNSON, 2000; WATERMAN et al., 2001; LAMB & FAUCHIER, 2001) die dargestellten, zu vermeidenden Fragetechniken/ -strategien zusammen (vgl. auch COOKE & DAVIS, 2001):

Tabelle 1-6: Unangemessene bzw. ungeeignete und problematische Fragestrategien

- suggestive Fragen, irreführende Fragen
- geschlossene Fragen wie Ja/Nein-Fragen, Entweder/Oder-Fragen, lenkende Fragen
- Wahlfragen
- Fragen, die eine kurze Antwort erfordern und
- wiederholte Fragen zu einem Aspekt
- beharrliches Einfordern von Namen, Zahlen, Daten und Zeiten
- komplexer Satzbau und kompliziertes Vokabular
- Verwendung von doppelten Verneinungen
- sofortiges Stellen der nächsten Frage
- Unterbrechen des Zeugen während dieser antwortet

Demnach erschien nach VOLBERT & STELLER (1997) die Strategie der „*Trichtertechnik*“ als die Methode der Wahl. Dabei wird durch entsprechenden Erzählanstoß zunächst ein freier Bericht evoziert. Im Rahmen eines *‘halbstrukturierten Interviews’* soll bei der nachfolgenden Befragung darauf geachtet werden, dass zuerst offen gefragt wird und erst spät eine spezifische Befragungsstrategie zum Einsatz kommt. Im Sinne einer Vermeidung von suggestiven Einflüssen ist dies unbedingt zu beachten.

Insgesamt ergeben die Befunde zur Forschung der Aussage- und Zeugenkompetenz intelligenzgeminderter Personen, dass der Umfang des freien Berichts oftmals eher spärlich ausfällt (MILNE, 1999), allerdings erscheint er meist sehr genau (vgl. HENRY & GUDJONSSON, 2003). Somit wird die Verwendung offener, allgemeiner Befragungstechniken von mehreren Seiten befürwortet (DENT, 1986; PERLMAN et al., 1994).

Nachstehende Tabelle fasst wiederum eher geeignete Befragungstechniken zusammen (vgl. auch COOKE & DAVIES, 2001):

Tabelle 1-7: Angemessene bzw. geeignete Befragungstechniken

- Rapport aufbauen
- freien Bericht anregen
- dann offene, allgemeine Fragen
- nur bedingt spezifische, geschlossene Fragen
- einfache und konkrete, nicht-abstrakte Formulierungen
- angepasstes Kommunikationsverhalten (einfacher Satzbau, verständliches Vokabular, langsames Sprechtempo, Vermeiden von Umgangssprache)
- geduldiges Abwarten im Falle von in verschiedenen Dimensionen verlangsamten Personen

Oftmals sind gerade intelligenzgeminderte Personen nicht in der Lage, ihre Aussage in einem ganzen, kontinuierlichen Interview zu machen, jedoch sehr wohl in mehreren kurzen Befragungen (COOKE & DAVIES, 2001). Bei der Planung und Durchführung von wiederholten Befragungen ist dies zu beachten, aber auch in Erwägung mit eventuell sich daraus ergebenden negativen oder sogar suggestiven (wiederholten) Befragungstechniken abzugleichen und einen auf die individuellen Bedürfnisse und Kompetenzen zugeschnittenen Mittelweg zu finden, um eine möglichst optimale Befragungssituation zu erreichen und das Maximum an Wahrheitsgehalt und Information zu gewinnen.

Zusammenfassend und abschließend sei angemerkt, dass sich die Grenzen einer optimalen Befragungssituation dort aufzeigen, wo man den Versuch unternimmt, dem Gegensatz zwischen spezifischerer Befragung (und damit Hinweisgebung), um einen vollständigen Bericht zu erhalten und Beeinflussbarkeit durch die Art der Hinweisgebung, gerecht zu werden.

COOKE & DAVIES (2001) ergänzen zudem noch einige mögliche Alternativen und Empfehlungen im übrigen Umgang mit lernbehinderten Zeugen:

- Wenn möglich sollte ein Besuch des Gerichtssaales stattfinden, um dem Zeugen einen Einblick in die Örtlichkeiten, Abläufe und beteiligte Personen zu geben.
- Zeugen sollten dazu ermutigt werden, Fragen zu stellen, falls sie weitere Erklärungen benötigen oder die Wiederholung einer gestellten Frage möglich ist. Zeugen sollten nicht das Gefühl haben, es wäre dumm oder unangebracht, wenn sie eine Frage (auch akustisch) nicht verstehen.
- Weiterhin sollten Zeugen darüber belehrt und informiert werden, dass das Zugeben einer Erinnerungslücke oder das Vergessen bestimmter Inhalte normal und verständlich ist und dies somit auch gesagt werden darf und soll, da dies in jedem Fall besser ist, als zu raten.

Die dargestellten Befunde sowie die forensische Praxis zeigen, dass gerade in Bezug auf die Vernehmung von Zeugen mit Intelligenzminderung noch Unsicherheit und vor allem Schulungsbedarf und intensive Vermittlung von entsprechend geeigneten Interviewtechniken und Hintergrundwissen notwendig bzw. von großer Bedeutung sind, um auch diesen Menschen eine möglichst optimale und ihren Bedürfnissen entsprechende Ausgangsvoraussetzung zu schaffen, eine zuverlässige und ihren Kompetenzen angepasste Zeugenaussage abzugeben.

2. Fragestellung

2.1 Fragen zur Untersuchung der Erinnerungsleistung und dem source- bzw. reality monitoring

Primäres Ziel und Gegenstand dieser Arbeit im Rahmen der Gedächtnis- und Zeugenpsychologie sind zum einen die Untersuchung von Erinnerungs- und Abrufleistungen aus dem Gedächtnis sowie von Quellenidentifikationsleistungen - also die Frage, inwieweit Personen mit einer Lernbehinderung in der Lage sind, internal generierte von external erhaltenen Informationen abzugrenzen bzw. unterschiedliche externale Quellen zu differenzieren. Auch die Analyse von weiteren Fehlerarten wie Konfabulationen, Intrusionen und sonstigen Erinnerungsfehlern bei Jugendlichen mit einer Intelligenzminderung sollen beleuchtet werden.

Darüber hinaus befasst sich die vorliegende Arbeit in einen längsschnittlichen Aspekt mit den Veränderungen von Gedächtnisleistungen über die Zeit hinweg, punktuell und im Vergleich zu verschiedenen Befragungszeitpunkten.

Auch werden die daraus gewonnen Befunde auf eine fragliche, gegenseitig voneinander abhängige Wechselwirkung hin untersucht.

Für die vorliegende Dissertation ergeben sich im Einzelnen folgende Fragestellungen:

2.1.1 Fragen zu den Unterschieden zwischen externaler 'source monitoring' - Gruppe und 'reality monitoring' - Gruppe

- Wie gut können Jugendliche mit Intelligenzminderung zwischen verschiedenen Informationsquellen differenzieren? Es ergeben sich diesbezüglich die weiteren, offenen Fragestellungen:
- Ist die Differenzierung zwischen zwei externalen Quellen nach einem Delay von einer Woche schwerer oder leichter als die Differenzierung zwischen einer externalen und einer internalen Quelle?
- Ist die Differenzierung zwischen zwei externalen Quellen nach einem Delay von sieben oder acht Wochen schwerer oder leichter als die Differenzierung zwischen einer externalen und einer internalen Quelle, wenn schon eine Befragung vorausgegangen ist?
- Ist die Differenzierung zwischen zwei externalen Quellen schwerer oder leichter als die Differenzierung zwischen einer externalen und einer internalen Quelle, wenn die erste Befragung nach einem Delay von sieben oder acht Wochen stattfindet?

Hier muss also statistisch gesehen von einer zweiseitigen Hypothesenprüfung ausgegangen werden.

Weiterhin ergibt sich hinsichtlich der zeitlichen Aspekte die Frage, welche Kurzzeit- und Langzeiteffekte sich in Bezug auf die Gedächtnisleistung bei Jugendlichen mit Intelligenzminderung zeigen.

2.1.2 Fragen zur Veränderung der Gedächtnisleistung über die Zeit

- Ist die Gedächtnisleistung der externalen 'source monitoring' - Gruppe zum zweiten Befragungszeitpunkt schlechter als zum ersten?
- Kann sich die externe 'source monitoring' - Gruppe zum zweiten Befragungszeitpunkt an eine der beiden Dia-Serien besser erinnern als an die andere?
- Ist die Gedächtnisleistung der 'reality monitoring' - Gruppe zum zweiten Befragungszeitpunkt schlechter als zum ersten?
- Kann sich die 'reality monitoring' - Gruppe zum zweiten Befragungszeitpunkt entweder an die Dia-Serie oder an die ausgedachte Geschichte besser erinnern?

Aus diesen Fragestellungen ergibt sich die allgemeine Alternativhypothese:

„Die Gedächtnisleistung beider Gruppen ist zum zweiten Befragungszeitpunkt schlechter als zum ersten Befragungszeitpunkt.“

Statistisch wird demnach einseitig getestet.

2.1.3 Fragen zum Zeitpunkt der Befragung

- Macht es für die externalen 'source monitoring' - Gruppe einen Unterschied, ob die erste Befragung nach einer oder nach ca. acht Wochen stattfindet?
- Macht es für die 'reality monitoring' - Gruppe einen Unterschied, ob die erste Befragung nach einer oder nach ca. acht Wochen stattfindet?
- Unterscheidet sich die Gedächtnisleistung der externalen 'source monitoring' - Gruppe nach einem Delay von ca. acht Wochen in Abhängigkeit davon, ob dieser Befragung schon ein Interview vorausgegangen ist oder nicht?

- Unterscheidet sich die Gedächtnisleistung der 'reality monitoring' – Gruppe nach einem Delay von ca. acht Wochen in Abhängigkeit davon, ob dieser Befragung schon ein Interview vorausgegangen ist oder nicht?

Aus diesen Fragestellungen ergeben sich folgende allgemeine Arbeitshypothesen:

„Die Gedächtnisleistung ist schlechter, wenn die erste Befragung bereits nach einer Woche und nicht erst nach acht Wochen durchgeführt wird.“

„Die Gedächtnisleistung zum zweiten Befragungszeitpunkt ist schlechter, wenn davor schon eine Befragung durchgeführt wurde.“

Statistisch wird demnach wiederum einseitig getestet.

2.1.3 Fragen zu Haupt- und Interaktionseffekten

- Gibt es zwischen den Versuchsgruppen zu den Befragungszeitpunkten T1 und T2 Unterschiede?
- Gibt es zwischen den Befragungszeitpunkten T1 und T2 Unterschiede für die Versuchsgruppen?
- Gibt es in Abhängigkeit vom Befragungszeitpunkt Unterschiede zwischen der externalen 'source monitoring' – Gruppe und der 'reality monitoring' – Gruppe?

Aus dieser Fragestellung ergeben sich die allgemeinen Arbeitshypothesen:

„Die Versuchsgruppen unterscheiden sich zu den beiden Befragungszeitpunkten nicht voneinander.“

„Für die beiden Versuchsgruppen lassen sich keine Unterschiede zwischen den Befragungszeitpunkten T1 und T2 feststellen.“

„Der Faktor Zeit verschlechtert die Gedächtnisleistung mehr, wenn zwischen zwei externalen Quellen differenziert werden muss als wenn zwischen einer externalen und einer internalen Quelle differenziert werden muss.“

2.1.5 sonstige Fragen

- Sind Jugendliche mit Intelligenzminderung in der Lage, vorgegebenes Reizmaterial mit selbst erdachten Inhalten anzureichern und zu ergänzen und diese Inhalte über längere Behaltensintervalle im Gedächtnis zu speichern?

Hier handelt es sich um eine offene Fragestellung, da zu diesem Punkt nicht auf fundierte empirische Ergebnisse zurückgegriffen werden kann. Die diesbezüglichen Hypothesen der Arbeit gehen davon aus, dass es den Jugendlichen entgegen den Ausführungen von GREUEL et al. (1998) dennoch möglich sein müsste, mit entsprechender Hilfestellung und Instruktion Inhalte zu erfinden. Es wird außerdem erwartet, dass sich diese Inhalte zumindest zum Teil auch als Intrusionen bei den Abfragen niederschlagen.

2.2 Fragen zur Glaubhaftigkeitsuntersuchung

2.2.1 allgemeine Fragen und Fragen zum generellen Differenzierungspotential der Aussagemerkmale

- Wie gut können Jugendliche mit einer Intelligenzminderung Falschaussagen bzw. Aussagen ohne Erlebnishintergrund im Vergleich zu wahren, erlebnisbasierten Aussagen grundsätzlich generieren sowie plausibel und glaubhaft darstellen?
- Wie gut sind die Jugendlichen auch über einen längeren (forensisch–praxisrelevanten) Zeitraum, d.h. zu zwei verschiedenen Befragungszeitpunkten in der Lage, das Wahrgenommene bzw. Erfundene zu erinnern?
- Zeigen sich hierbei Unterschiede zwischen den erlebnisbasierten Aussagen und den selbst ausgedachten?
- Welche Aussagemerkmale erweisen sich als Trennungssindikator zwischen erlebnisbasierten und erfundenen Aussagen?

2.2.2 Fragen hinsichtlich verschiedener Befragungszeitpunkte und des Delays

- Inwieweit trennen die jeweiligen Aussagemerkmale zu zwei unterschiedlichen Befragungszeitpunkten in Zeugenaussagen von Jugendlichen mit einer Intelligenzminderung?
- In welche Richtung gehen die eventuell gefundenen Unterschiede zwischen den beiden Befragungszeitpunkten?

- Gibt es Aussagemerkmale, die zum ersten Befragungszeitpunkt T1 zwischen den Versuchsgruppen differenzieren und/oder zum zweiten Befragungszeitpunkt T2 nicht oder umgekehrt?

2.2.3 Fragen hinsichtlich der gutachterlichen, forensischen, polizeilichen und richterlichen Praxis

- In diesem Zusammenhang sind vor allem Aspekte - gerade bei Menschen mit intellektuellen Einschränkungen – von geeigneten Befragungs- und Interviewtechniken von zentraler Bedeutung im Rahmen der Begutachtung von Zeugen wie auch in Situationen vor Gericht bzw. anderen (autoritären) Institutionen oder Personen, um trotz kognitiver Einschränkungen und anderer (sozialer, motivationaler oder emotionaler) Einflussfaktoren möglichst wahrheitsgemäße, zuverlässige, vollständige und umfassende Aussagen und Schilderungen zu ermöglichen und zu erhalten.
- Weiterhin soll der Einfluss falscher oder irreführender Informationen, anderer Formen suggestiver Beeinflussung, kommunikativer Bedingungen und sonstiger Faktoren vor, während und nach dem relevanten (beobachteten) Ereignis zumindest ansatzweise beleuchtet werden.
- Zudem stellt sich die Frage, inwieweit interdisziplinäre und angewandte Forschung und Zusammenarbeit im Bereich der forensischen Psychologie und Psychologie der Zeugenaussage Implikationen und weitere Fortschritte für eine Optimierung im Umgang mit behinderten Menschen - vor allem als Zeugen und Opfer von Straftaten gegen die sexuelle Selbstbestimmung, jedoch auch im Sinne eines dringend gebotenen und aktuell immer wieder diskutierten Opferschutzes - darstellen können.

2.2.4 Fragen hinsichtlich der Anwendbarkeit auf allgemeinspsychologische Wahrnehmungs- und Gedächtnisprozesse

- Lässt sich der Kriterienkatalog, der vor allem für den Straftatbestand des sexuellen Missbrauchs entwickelt wurde auf allgemeinspsychologische Wahrnehmungs- und Gedächtnisprozesse und die Zeugenpsychologie anwenden, besonders im speziellen Fall der Intelligenzminderung bei Jugendlichen ?

3. Methoden

Im Folgenden werden die methodischen Aspekte und Vorgehensweisen für die beiden Untersuchungen jeweils getrennt voneinander dargestellt. Begonnen wird mit der Untersuchung zur Gedächtnisleistung (recall) bzw. zum 'source-' / 'reality-monitoring'.

3.1 Untersuchung zur Erinnerungsleistung und dem source- bzw. reality monitoring

3.1.1 Versuchspersonen

An dieser ersten von zwei Untersuchungen der vorliegenden Arbeit nahmen als Versuchspersonen insgesamt $N = 44$ Jugendliche und jugendliche Erwachsene mit einer Intelligenzminderung im Alter zwischen 16 und 19 Jahren teil ($MW = 17,95$; $SD = 0,75$ zum Zeitpunkt der Präsentation des Versuchsmaterials).

Die Jugendlichen waren allesamt als Schüler bzw. Auszubildende im Berufsbildungswerk (B.B.W.) St. Franziskus in Abensberg in heilpädagogischen Jugendwohngruppen untergebracht. Das B.B.W. Abensberg versteht sich als Zentrum für berufliche Bildung und Rehabilitation mit einer integrierten Berufsschule zur sonderpädagogischen und individuellen Lernförderung (siehe auch <http://www bbw-abensberg.de>).

Um die spätere Vergleichbarkeit der gewonnenen Daten zu gewährleisten, wurde festgelegt, dass die folgenden Personengruppen *nicht* in die Stichprobe miteinbezogen werden sollten:

- Jugendliche, deren Wahrnehmungsfähigkeiten auf einem oder mehreren Sinneskanälen eingeschränkt sind (Taubheit, Blindheit)
- Jugendliche mit expliziten Sprachstörungen (z. B. Aphasien)
- Jugendliche mit psychiatrischen Störungen, die mit Wahrnehmungsverzerrungen einhergehen können (z. B. Borderline – Patienten)

Die genannten Kriterien stellen also im Wesentlichen Ausschlusskriterien dar.

Bezüglich des Intelligenzniveaus der Versuchspersonen bzw. der Schwere der Intelligenzminderung muss angemerkt werden, dass eine entsprechende Auswahl von Versuchspersonen – insbesondere respektive der oben genannten Ausschlusskriterien nicht ganz einfach war. Aus den Akten konnten neben dem testpsychologisch erhobenen Intelligenzniveau Hinweise auf Störungen entnommen werden. Da die vorliegenden Störungsbilder von Jugendlichem zu Jugendlichem stark variieren, kann keine umfassende Darstellung vorgenommen werden. Die Defizite liegen in den unterschiedlichsten psychischen und körperlichen Systemen.

Bei einigen der Jugendlichen war nicht unbedingt eine Intelligenzminderung im eigentlichen Sinn (IQ-Wert) vorhanden, trotzdem diese nach Auskunft der Psychologen vor Ort aufgrund von Auffälligkeiten im sozialen Verhalten (Aggression, dissoziales Verhalten und Ähnliches) sowie im Bereich Aufmerksamkeitsstörungen/Aufmerksamkeitsdefizite, Konzentrationsstörungen und Entwicklungsverzögerung Probleme hatten. Auf die Schwierigkeiten der Auswahl geeigneter Versuchspersonen im Rahmen solcher Untersuchungen soll im Kapitel 'Diskussion' näher eingegangen werden bzw. wird auf die bereits einleitend dargestellten theoretischen und empirischen Befunde verwiesen.

Letztendlich ergibt sich bezüglich dieser ersten Teiluntersuchung folgendes Bild hinsichtlich der IQ-Werte:

Abbildung 3-1: Verteilung der IQ-Werte (source- und reality monitoring)

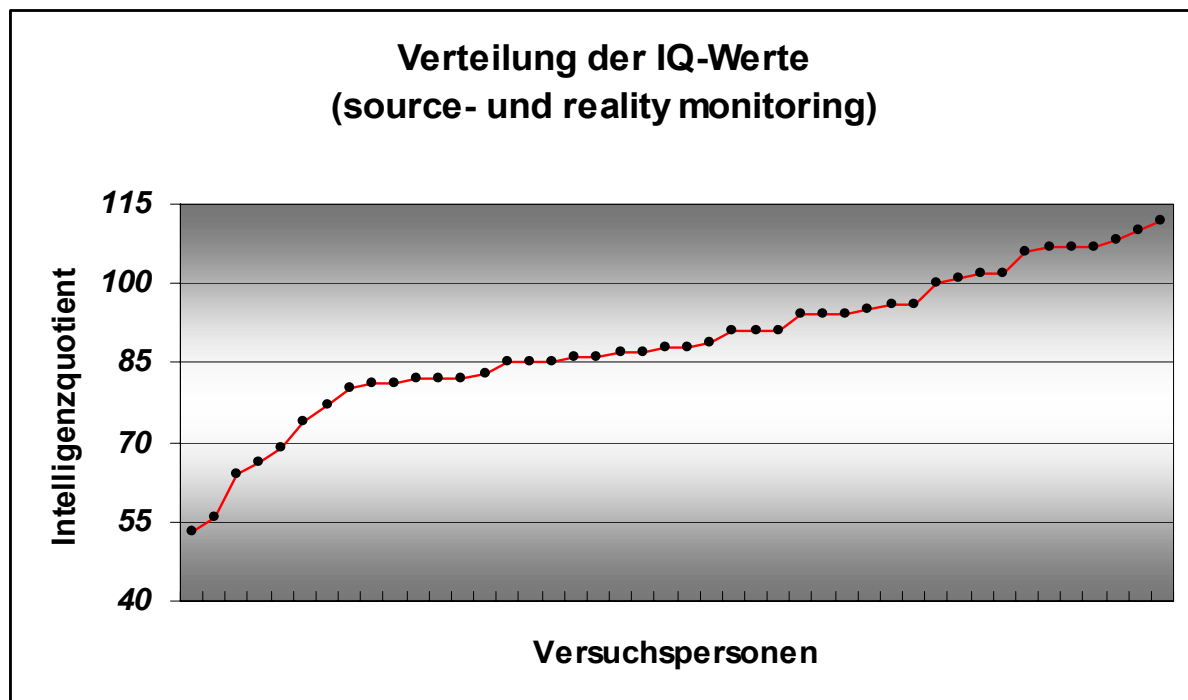


Tabelle 3-1: IQ-Daten 1 (Untersuchung 'source-' und 'reality monitoring')

N	Mittelwert	SD	Median	Minimum	Maximum	Spannweite
44	88,64	13,80	88,00	53	112	59

Insgesamt kann somit von einem relativ hohen Intelligenzniveau ausgegangen werden – im Rahmen der Population 'Menschen mit Intelligenzminderung oder Lernbehinderung'.

Die Leitung des B.B.W. Abensberg und deren Mitarbeiter erklärten sich dankenswerterweise bereit, bei der Einwerbung der geeigneten Versuchspersonen zu helfen. Hierzu wurde ein Schreiben mit einer kurzen Darstellung des Projektes und dessen grobem Ablauf an die

Eltern, Erziehungsberechtigten oder anderweitig verantwortliche Personen geschickt und um die Erlaubnis an der Teilnahme der Untersuchung ersucht. Von den ca. 120 angeschriebenen Erziehungsberechtigten erklärten sich letztendlich 48 bereit, die bzw. den Jugendlichen teilnehmen zu lassen – das Einverständnis des Jugendlichen selbst natürlich ebenfalls vorausgesetzt. Wie auch von Seiten der psychologischen Mitarbeiter des B.B.W. vermutet, war der Rücklauf der Anschreiben vor allem aufgrund der oft problematischen und ungeordneten Familienverhältnisse relativ gering. Des Weiteren reduzierte sich aus verschiedenen Gründen die Teilnehmerzahl nochmals um vier auf schließlich 44. Davon waren 25 männlich und 19 weiblich.

Jede der Versuchspersonen erhielt als Dankeschön für die Teilnahme einen attraktiven Kinogutschein (freie Film- und Platzwahl) im Wert von ca. acht Euro.

3.1.2 Versuchsmaterial

Als Versuchsmaterial wurden in dieser Untersuchung zum einen zwei Dia-Serien eines gestellten Verkehrsunfalls, zum anderen als Informationsgrundlage für die Befragungen eine zur der zweiten Dia-Serie analoge Darstellung, die von den Versuchspersonen selbst erfunden wurde. Dies wird im Folgenden noch genauer erläutert.

Die beiden Dia-Serien wurden aus insgesamt sechs Dia-Serien ausgewählt, die dem Lehrstuhl bereits vorlagen, da sie in einer Diplomarbeit als Stimulusmaterial verwendet wurden und auch für diesen Zweck produziert wurden. Die sich an einer Kreuzung im Regensburger Umland abspielenden Szenen zeigen jeweils die realistisch wirkende Nachstellung von leichten Verkehrsunfällen mit der Beteiligung verschiedener Personen und Fahrzeugen. Die Erlaubnis, die Szenen für wissenschaftliche Zwecke zu verwenden, lag von allen Beteiligten vor.

Bei der Auswahl der beiden Dia-Serien wurde darauf geachtet, dass sich die dabei vorkommenden Fahrzeuge und Personen soweit voneinander unterschieden, dass es für die Versuchspersonen weder zu leicht noch zu schwer war, sich an Abläufe, Personen und andere Details zu erinnern.

Die Dia-Serien bestanden jeweils aus 21 einzelnen Szenen, wobei der Wechsel von Bild zu Bild automatisch nach einer Darbietungszeit von 10 Sekunden erfolgte. Das erste bzw. letzte Dia zeigte jeweils den Beginn bzw. das Ende der Vorführung an.

Die Dia-Serien finden sich in der Druckversion als auch in der als elektronisches Medium veröffentlichten Version dieser Dissertation in *Anhang D1-D2* bzw. auch unter (IrfanView):

http://www.uni-regensburg.de/Fakultaeten/phil_Fak_II/Psychologie/Doktoranden/absolventen/berger_oli/index.php

3.1.2.1 Der Inhalt der Dia-Serien

Die Inhalte bzw. Abläufe der beiden Dia-Serien lassen sich zusammenfassend in etwa folgendermaßen wiedergeben (siehe auch *Anhang D1 und D2*):

3.1.2.1.1 Dia-Serie 1 – PKW-PKW

An einer Y – förmigen Kreuzung kommt es zwischen einem silbernen Ford Focus, der rückwärts aus einer linken Abzweigung herausfährt, und einem roten Opel Corsa zu einem leichten Auffahrunfall bzw. Zusammenstoß. Der silberne Ford Focus wird von einer blonden Frau in einem schwarzen Minikleid gefahren, der rote Opel Corsa von einem Mann, der mit einem weißen Hemd und einer schwarzen Jeans bekleidet ist. Als Zeugen beobachten den Unfall eine Radfahrerin in einem schwarzen Kleid und eine Frau, die ihr Kind in einem Kinderwagen schiebt und eine Jeans und T – Shirt trägt. Das Alter der Personen bewegt sich zwischen etwa 20 und 35 Jahren. Per Zettel und per Handy tauschen alle Beteiligten Adressen und Telefonnummern aus. Die beiden direkt am Unfall Beteiligten verabschieden sich anschließend per Handschlag.

3.1.2.1.2 Dia-Serie 2 – PKW-Fahrrad

An derselben Y – förmigen Kreuzung wie in Dia-Serie 1 kommt es zu einem Unfall in Form eines Zusammenstoßes zwischen einer Frau, die einen grünen Opel Corsa fährt und einem männlichen Radfahrer, welcher aus der besagten Abzweigung nach links in Richtung des PKW abbiegen will. Der Radfahrer ist mit einem gemusterten Hemd und einer Jeans bekleidet, die Autofahrerin mit Jeans und grauem T-Shirt. Als Zeugen beobachten den Unfall eine Frau mit Kinderwagen, die eine weiße ärmellose Bluse trägt und eine Frau am rechten Straßenrand, die ihr Auto aufsperrt und mit einem gelben T – Shirt und einem Rock bekleidet ist. Das Alter der beteiligten Personen liegt wie in Dia-Serie 1 ca. zwischen 20 und 35. Ähnlich wie in Dia-Serie 1 werden Adressen ausgetauscht, diesmal jedoch ohne Handy. Der offensichtlich unverletzte Radfahrer und die junge Autofahrerin verabschieden sich ebenso per Handschlag.

Dabei kommt keine der beteiligten Personen in der jeweils anderen Dia-Serie ebenfalls vor.

3.1.2.2 Die selbst erfundene Unfall-Serie – analog Dia-Serie ‘PKW-Fahrrad’

Der Hälfte der Versuchspersonen (N = 22) wurde lediglich eine Dia-Serie, nämlich Dia-Serie 1 (PKW-PKW) präsentiert. Anstelle der zweiten Dia-Serie 2 (PKW-Fahrrad) sollte sich diese Versuchspersonengruppe selbst einen weiteren Unfall ausdenken, wobei hier einige Vorgaben und Anhaltspunkte gegeben wurden, an denen sich orientiert werden sollte. Dieser Unfall sollte sich zwar an derselben Straßenkreuzung ereignen wie der Unfall aus Dia-Serie 1 mit ebenfalls einem männlichen Radfahrer und einer weiblichen Autofahrerin. Der Unfall sollte wiederum von zwei weiblichen Zeugen beobachtet werden. Jedoch wurden die Jugendlichen entsprechend angewiesen, nicht einfach die Situation aus der ersten Dia-Serie zu übernehmen, sondern evtl. auch hinsichtlich der Richtungsangaben und der Schuldfrage zu variieren.

Des Weiteren wurden die Versuchspersonen instruiert, den Unfallhergang schriftlich und evtl. anhand einer Skizze zu beschreiben und mittels eines ausgegebenen Bogens in einer vorgegebenen Tabelle die beteiligten Personen und Fahrzeuge exakt zu beschreiben.

Die genauen Instruktionen sowie der auszufüllende Bogen finden sich in *Anhang A*.

3.1.2.3 Präsentation des Versuchsmaterials und Instruktion der Versuchspersonen

Wie folgt wurde den beiden Versuchsgruppen wurde am gleichen Tag jeweils hintereinander das Stimulusmaterial präsentiert sowie erhielten sie die entsprechenden Instruktionen:

3.1.2.3.1 externe ‘source monitoring’ - Gruppe (ESM - Gruppe)

Die Versuchspersonen (n = 22) sahen per Video-Beamer zunächst Dia-Serie 1 (PKW-PKW), im Anschluss daran Dia-Serie 2 (PKW-Fahrrad).

Alle Jugendlichen wurden gebeten, sich das Gesehene gut zu merken, da sie in den kommenden Wochen dazu befragt werden sollten. Eine Hälfte der Gruppe würde dabei zweimal befragt werden – das erste Mal in einer Woche ab heute, das zweite Mal in ca. sieben oder acht Wochen. Die andere Hälfte müsse jedoch nur einmal interviewt werden und zwar ebenfalls in ca. acht Wochen. Danach verließen die Jugendlichen den Raum.

3.1.2.3.2 'reality monitoring' – Gruppe (RM – Gruppe)

Im direkten Anschluss an die ESM-Gruppe wurde im selben Raum den Versuchspersonen der RM-Gruppe (n = 22) zunächst wiederum per Video-Beamer ebenfalls Dia-Serie 1 (PKW-PKW) vorgeführt. Darauf hin wurden Papier bzw. die erwähnten Bögen und Stifte verteilt und die Jugendlichen instruiert und gebeten, sich nun – wie oben in Kapitel 3.1.2.2 erläutert – anhand der Vorgaben selbst einen Unfall auszudenken und diesen schriftlich festzuhalten. Dazu hatten die Versuchspersonen ca. 30 Minuten Zeit. Nachdem die Bögen wieder eingesammelt waren, erfolgte analog zur externalen 'source monitoring' - Gruppe dieselbe Instruktion bezüglich der weiteren Befragungsmodalitäten und -zeitpunkte für je die Hälfte der Versuchspersonen.

Über den Zweck der Untersuchung wurden die Versuchspersonen dergestalt informiert, als dass ihnen gesagt wurde, dass es sich hierbei um ein Gedächtnisexperiment im Rahmen von Zeugenaussagen handeln würde. Ähnlich wie bei tatsächlichen Gerichtsverhandlungen und im wirklichen Leben solle dabei untersucht werden, wie viel bzw. was die Versuchspersonen nach einer bestimmten Zeit noch von den gesehenen Unfällen wissen. Somit war die Aufgabe mit einem intentionalen Lernvorgang zu vergleichen, da die Versuchspersonen wussten, dass es darauf ankam, sich die Dia-Serien bzw. den eigenen selbst ausgedachten und beschriebenen Unfallhergang zu merken.

3.1.3 Experimentalplan

Der nachstehende Experimentalplan verdeutlicht und ergänzt die Modalitäten hinsichtlich Versuchsmaterial, Versuchsbedingungen, Gruppengrößen und Befragungszeitpunkten.

Tabelle 3-2: Experimentalplan (1) - 'source-/ reality monitoring'

Gruppe	Versuchsmaterial			Befragungszeitpunkt		Zahl d. VPn	
	<i>Dia-Serie 1 (PKW- PKW)</i>	<i>Dia-Serie 2 (PKW- Fahrrad)</i>	<i>Serie 2 (erfunden)</i>	<i>T1</i>	<i>T2</i>	<i>n</i>	<i>N</i>
ESM	X	X		X	X	11	22
	X	X			X	11	
RM	X		X	X	X	12	22
	X		X		X	10	

Auf der Basis dieses Experimentalplans sollte den folgenden grundsätzlichen Fragestellungen nachgegangen werden:

- Gibt es Unterschiede in der Gedächtnisleistung zwischen den Versuchsgruppen externes 'source monitoring' und 'reality monitoring'?
- Wie verändert sich die Zeugenkompetenz über die Zeit?
- Wie unterscheidet sich die Gedächtnisleistung zu den verschiedenen Befragungszeitpunkten?
- Bestehen hinsichtlich der Gedächtnisleistung Wechselwirkungen zwischen Gruppenzugehörigkeit und Befragungszeitpunkt?

Da neben den kontrollierten unabhängigen Variablen auch andere mögliche Einflussgrößen (Delay der Befragung, Berufs- bzw. Ausbildungszweig, Geschlecht der Versuchspersonen) vorkommen, wurde auch dank und mithilfe einer sehr engagierten Mitarbeiterin des B.B.W. versucht, einen potentiellen Effekt der genannten Moderatorvariablen im - innerhalb des organisatorischen und logistischen Ablauf - größtmöglichen Ausmaß zu vermeiden bzw. zu kontrollieren. Somit war es ein wichtiges Anliegen, die Versuchsgruppen so gestalten, dass sie soweit wie möglich 'parallelisiert' bzw. ausbalanciert sind.

Dazu wurde zum einen versucht, den Abstand zwischen der Präsentation des Versuchsmaterials und den einzelnen Befragungszeitpunkten für jede Versuchsperson möglichst gleich zu halten. Die Reihenfolge der Befragung der Jugendlichen pro Versuchsgruppe wurde also über die Tage und Wochen der Interviewdurchführungen weitestgehend eingehalten.

Zum andern wurden den (Teil-)Versuchsgruppen die Versuchspersonen so zugeordnet, dass ebenso eine homogene Verteilung der verschiedenen Berufs- und Ausbildungszweige erreicht wurde.

Des Weiteren wurde bei der Verteilung der Versuchspersonen auf die Versuchsbedingungen darauf geachtet, dass ein relativ ausgewogenes Verhältnis bezüglich des Geschlechts vorlag.

Dies soll *Tabelle 3-3* verdeutlichen:

Tabelle 3-3: Geschlechtsspezifische Verteilung der Versuchspersonen auf die Versuchsgruppen ('source-/ reality monitoring')

Gruppe	♀		♂		Zahl der VPn	
	Befragungszeitpunkt		Befragungszeitpunkt			
	T1/T2	NT2	T1/T2	NT2	n	N
ESM	4	6	7	5	22	44
RM	5	4	7	6	22	
Zahl der VPn						
n ₁	9	10	14	11	44	
n ₂	19		25		44	
N	44					

NT2: Zeitpunkt, zu dem einige Versuchspersonen, das erste und einzige Mal befragt wurden.
(bezüglich des Delays identisch mit T2)

3.1.4 Durchführung der Befragungen

Aufgrund eines im Rahmen der Diplomarbeit zu einem Thema der Aussage- und Gedächtnispsychologie durchgeführten intensiven Interviewertrainings sowie seiner ca. einjährigen Tätigkeit als Sachverständiger für Glaubhaftigkeitsbegutachtung wurde von einer entsprechenden Interviewerkompetenz des Verfassers dieser Arbeit ausgegangen und auf ein explizites Interviewertraining verzichtet.

Dem aktuellen Forschungsstand sowie der forensischen und sachverständigen Praxis entsprechend wurde dennoch besonders auf die richtige Art der Fragestellung, den Aufbau von Rapport und einer im Übrigen möglichst optimalen Befragungssituation geachtet. Dabei sollten – aufbauend auf GREUEL et al. (1998) – vor allem offene Fragen, Wahlfragen, Leerfragen und Konträrfragen verwendet und aber auch in der weiteren Befragung Suggestivfragen, Erwartungsfragen, Voraussetzungsfragen und Ja-Nein-Fragen vermieden werden. Dennoch galt als Orientierung hierzu ein sich in *Anhang B* befindlicher Interviewleitfaden.

3.1.4.1 Zeitlicher Abstand der Befragungen zur Stimuluspräsentation

Genau eine Woche nach Präsentation des Versuchsmaterials wurde in einem dazu zur Verfügung gestellten Büro des B.B.W. Abensberg begonnen, die jeweiligen Interviews mit den Versuchspersonen einzeln durchzuführen. Die gesamte Dauer der Befragungsdurchfüh-

rung umfasste zum ersten Befragungszeitpunkt T1 nach einer Woche für beide Versuchsgruppen ($n = 23$) sieben Tage.

Genau sieben Wochen nach Präsentation des Versuchsmaterials wurde mit der Interviewdurchführung zum zweiten Befragungszeitpunkt T2 und NT2 begonnen ($N = 44$). Dabei wurden zusätzlich nun auch diejenigen Versuchspersonen ($n = 21$) interviewt, die an der ersten Befragung nicht teilgenommen hatten (NT2). Die Gesamtdauer dieses Befragungszeitraums umfasste ca. 13 Tage.

Dabei wurde – wie schon erwähnt – bei der zweiten Befragung T2 darauf geachtet, dass die Reihenfolge der zu befragenden Versuchspersonen tagesbezogen die selbe war, wie zu T1, alle Jugendlichen, die an zwei Befragungen teilnahmen also in etwa den selben zeitlichen Abstand zwischen ihrer ersten und der zweiten Befragung hatten.

3.1.4.2 Ablauf der Befragungen

In enger Kooperation und Organisation mit den verschiedenen Ausbildern und Mitarbeitern wurden die Jugendlichen einzeln zum Befragungsraum bestellt und dort interviewt.

Der konkrete Ablauf der einzelnen Befragungen lief zu allen Befragungszeitpunkten (T1, T2 und NT2) identisch nach dem Schema eines halbstrukturierten Interviews ab. Dabei saßen sich der/die Jugendliche und der Versuchsleiter an einem Tisch gegenüber. Sämtliche Interviews wurden mittels eines Konferenzmikrofons und eines Diktiergeräts aufgezeichnet. Diese Aufnahmen dienten dem Versuchsleiter später zum Anfertigen der Transkripte der Interviews. Parallel wurden zum Zwecke der Struktur der weiteren, differenzierten Befragung und um eventuelle unabsichtliche Suggestiv- oder Wiederholungsfragen zu vermeiden, stichpunktartig die von der Versuchsperson genannten Aspekte mitnotiert.

Die Dauer eines einzelnen Interviews betrug ca. zwischen 45 und 60 Minuten.

Inhaltlich folgte nach einer Vorphase des Interviews - der Begrüßung und des Herstellens von Rapports, wo man sich kurz (zum Teil sehr nett und lustig) über verschiedene, den/die Jugendlichen interessierende Themen unterhielt - die erste Phase des eigentlichen Interviews, welche im 'free recall' bestand. Hier sollte die Versuchspersonen - ohne vom Versuchsleiter durch eine Frage unterbrochen worden zu sein, sondern lediglich durch aktives Zuhören unterstützt - zunächst all das berichten, an das sie sich noch erinnern konnten. Traten hier (bereits) Schwierigkeiten auf, wurde einmalig versucht, den freien Bericht weiter anzuregen, indem nach weiteren Aspekten oder Details gefragt wurde. Kam dort auch nichts mehr nach, wurde zur zweiten Interviewphase, dem konkreten Nachfragen übergeleitet.

Wie aus dem im Anhang befindlichen Interviewleitfaden genauer ersichtlich, sollten hier sich aus dem freien Bericht evtl. ergebene Unklarheiten beseitigt werden und weitere, detaillierte Aspekte und Informationen zu den verschiedenen Bereichen (Personen, Unfallhergang, Örtlichkeiten, andere Umstände, ...) gewonnen werden.

Am Ende ihrer Ausführungen hatte jede Versuchsperson nochmals die Möglichkeit, all das zu ergänzen, was ihr noch zu irgendeinem Bereich einfiel.

Nach Verabschiedung und Aushändigung des Kinogutscheins wurde die Versuchsperson entlassen.

3.1.5 Auswertung der Befragungen

3.1.5.1 Transkription der Interviews

Fasst man die Versuchsgruppen und alle Befragungszeitpunkte zusammen, so stellt dies einen Gesamtumfang von 67 durchgeführten Interviews dar. Daraus wurden vom Versuchsleiter entsprechend 67 wortwörtliche Transkripte angefertigt.

3.1.5.2 Auswertung der Transkripte

Für beide Dia-Serien bzw. eine Dia-Serie und der selbst ausgedachten Unfall-Serie wurden auf Basis der Transkripte für alle Versuchspersonen die korrekt berichteten Details (im freien Bericht und auch auf Nachfrage) sowie die Fehler und einzelnen Fehlerarten mittels eines Auswerte-Schemas (siehe *Anhang C*) eingetragen und ausgezählt.

Auswertung der Details:

In Anlehnung an eine bereits verfasste Diplomarbeit (Kinshofer, 2002) wurde die korrekt berichteten Details anhand folgender Regeln ausgezählt, die jedoch für die vorliegende Dissertation überarbeitet und weiterentwickelt wurden:

1. Jedes sinnvolle Detail wird mit einem Punkt bewertet.

Beispiel: „Im Auto war ein Mann und auf dem Rad eine Frau.“

Bewertung: Auto – Mann – Rad – Frau

⇒ 4 Detailpunkte

- 2. Ein sinnvolles Detail muss nicht zwingend aus einem Wort bestehen. Feststehende Begriffe, die sinngemäß zusammengehören, werden als eine Einheit zusammengefasst und als ein Detail gewertet.**

Beispiel: „Das Auto bog nach links ab.“

Bewertung: Auto – bog nach links ab ⇒ 2 Detailpunkte

- 3. Jedes Detail erhält grundsätzlich nur einen Punkt. Wird es im Verlauf des Interviews mehrmals erwähnt, so wird nur die erste Benennung bewertet.**

Beispiel: „Der Zeuge war ein Mann.....Der Zeuge hatte ein Rad.“

Bewertung: Zeuge – Mann – Rad ⇒ 3 Detailpunkte

- 4. Werden im Verlauf des Interviews Konkretisierungen von einem Detail vorgenommen, so werden nur diese mit einem Punkt bewertet.**

Beispiel: „Das Oberteil war rötlich....es war dunkelrot.“

Bewertung: Oberteil – dunkelrot ⇒ 2 Detailpunkte

- 5. Zahlenmäßige Größenangaben erhalten nur einen Punkt, wenn sie konkret sind.**

Beispiel 1: „Der Mann war 1,80.“

Bewertung: Mann – 1,80 ⇒ 2 Detailpunkte

Beispiel 2: „Die Frau war zwischen 1,70 und 1,80.“

Bewertung: Frau ⇒ 1 Detailpunkt

- 6. Negationen erhalten keinen extra Punkt.**

Beispiel 1: „Der Mann hatte ein Handy.“

Bewertung: Mann – Handy ⇒ 2 Detailpunkte

Beispiel 2: „Der Mann hatte kein Handy.“

Bewertung: Mann – kein Handy ⇒ 2 Detailpunkte

- 7. Ein Adjektiv erhält nur dann einen Punkt, wenn es eine zusätzliche Information liefert, die zudem objektiv ist.**

Beispiel 1: „Das Auto ist grün.“

Bewertung: Auto – grün ⇒ 2 Detailpunkte

Beispiel 2: „Die Frau trug eine schöne Jacke.“

Bewertung: Frau – Jacke ⇒ 2 Detailpunkte

8. Zusammengesetzte Wörter erhalten dann einen Punkt, wenn sie eine zusätzliche Information liefern.

Beispiel: „Sonnenbrille“

Bewertung: Sonnen – brille

⇒ 2 Detailpunkte

9. Für allgemeine Verben, die keine konkrete Ausprägung widerspiegeln, wird kein Punkt vergeben. Nur konkrete Ausprägungen von Verben erhalten einen Punkt.

Beispiel 1: „fahren“

⇒ 0 Detailpunkte

Beispiel 2: „rasen“

⇒ 1 Detailpunkt

10. Unbestimmte Zeitangaben werden nicht bewertet, weil sie keine zusätzliche Information liefern.

Beispiel: „danach“

⇒ 0 Detailpunkte

Die Details wurden zu jedem der Befragungszeitpunkte getrennt für die beiden Versuchsmaterialien ausgezählt. Dabei wurde jeweils noch zwischen den korrekten Details aus den freien Berichten und den korrekten Details, die auf die konkreten Fragen berichtet wurden, unterschieden. Die Summe dieser beide Kategorien ergab dann die Gesamtanzahl der korrekten Details in einem Interview (korrekte Details gesamt).

Auswertung der Fehler

Die Gesamtfehlerzahl ergab sich aus der Summe der in die verschiedenen Arten aufgeteilten Fehler (Quellenverwechslungen, schematypische Intrusionen, Konfabulationen und allgemeine Erinnerungsfehler).

Innerhalb der Quellenverwechslungen wurde weiterhin insofern differenziert, als dass unter „Quellenverwechslungen innerhalb“ (QueVw intra) diejenigen Fehler fielen, die von einer Versuchsperson begangen wurden, wenn sie eine Beobachtung innerhalb eines Versuchsmaterials falsch zuordnete; als „Quellenverwechslungen zwischen“ (QueVw inter) wurde ein Fehler gezählt, wenn eine Versuchsperson ein beobachtetes Detail aus Dia-Serie 1 (PKW-PKW) der zweiten Dia-Serie (PKW-Fahrrad) bzw. der selbst ausgedachten Unfall-Serie zuordnete oder umgekehrt.

Ein Fehler galt dann als schematypische Intrusion, wenn von etwas berichtet wurde, das objektiv in keinem der Versuchsmaterialien auftrat, jedoch dem allgemeinen Schema von „Unfall“ zugewiesen werden konnte.

Unter Konfabulationen fielen all diejenigen Äußerungen, Schilderungen oder Elemente der Versuchspersonen, die zweifelsfrei von ihnen erfunden waren und auf keinem der Versuchsmaterialien basierten.

Die Kategorie „Allgemeine Erinnerungsfehler“ diente dazu, Aussagen wie z.B. eine falsche Beschreibung der Kleiderfarbe, eine unzutreffende Marke des Autos oder eine unzutreffende Größenangabe einzuordnen.

Alle Fehlerarten sowie die Gesamtfehlerzahl wurden analog zu den korrekt berichteten Details zu jedem Befragungszeitpunkt getrennt nach den beiden Versuchsmaterialien ausgezählt. Dabei ging es nicht um das bloße Zählen von Wörtern, sondern es wurden Wörter sinngemäß zu einer Einheit zusammengefasst und dann der jeweiligen Fehlerart zugewiesen.

3.1.6 Variablenübersicht

Als jeweils unabhängige Variablen wurden in der vorliegenden Teiluntersuchung die verschiedenen Befragungszeitpunkte (T1/T2 vs. NT2) und die beiden Versuchsbedingungen (external - external vs. external - internal) systematisch variiert. Daraus ergibt sich ein 2x2 – faktorielles Design mit insgesamt 10 intervallskalierten, abhängigen Variablen:

Tabelle 3-4: Abhängige Variablen (1) - 'source-/ 'reality-monitoring'

Kategorie	abhängige Variable
korrekt berichtete Details	• korrekte Details 'gesamt'
	• korrekte Details im 'free recall'
	• korrekte Details 'auf Nachfrage'
Fehler	• Fehler 'gesamt'
	• Quellenverwechslungen 'gesamt'
	• Quellenverwechslungen 'innerhalb' der Dia- bzw. Unfall-Serie
	• Quellenverwechslungen 'zwischen' den Serien
	• schematypische Intrusionen
	• Konfabulationen
	• allgemeine Erinnerungsfehler

3.1.7 Statistische Auswertung

Alle statistischen Analysen wurden anhand des Statistikprogramms 'SPSS für Windows - Version 12.0' durchgeführt; Abbildungen, Diagramme und graphische Darstellungen mittels Microsoft Office Excel 2003.

3.1.7.1 Signifikanztests

Für alle aus den Interviews gewonnenen abhängigen Variablen wurde von intervallskalierten Daten ausgegangen. Ebenfalls wurde in jeder Subgruppe dieser Teiluntersuchung für jede dieser Variablen anhand des Einstichproben-KOLMOGOROV-SMIRNOV-Test bzw. des LEVENE-Test überprüft, welche Verteilungsform bzw. ob Varianzhomogenität besteht oder nicht. Der LEVENE-Test ist dabei selbst nicht von einer Normalverteilung der Grundgesamtheit abhängig (vgl. JANSSEN & LATZ, 1999).

Für alle vier Gruppen konnte schließlich von einer Normalverteilungsannahme ausgegangen werden.

Somit konnten zur Überprüfung der Mittelwertsunterschiede verschiedene varianzanalytische Verfahren verwendet werden.

Im Einzelnen erwiesen sich die folgenden als hierfür geeignet:

- a) T – Tests für abhängige und unabhängige Stichproben
- b) Allgemeines Lineares Modell ('ALM')

Die nach BORTZ (1999) zur Verwendung des T-Tests für abhängige Stichproben geforderte Voraussetzung einer positiven Kovarianz (Korrelation) zweier Messwertreihen war in fast allen Fällen für fast alle Variablen erfüllt (54 von 60 untersuchten abhängigen Variablen), jedoch ergaben diese Zusammenhänge bei einer angenommenen Irrtumswahrscheinlichkeit von 5 % und einer zweiseitigen Testung nicht immer ein signifikantes Ergebnis.

Der T-Test für abhängige als auch für unabhängige Stichproben erweist sich insgesamt allerdings als relativ robust gegenüber eventuellen Verletzungen der geforderten Voraussetzungen (vgl. BORTZ, 1999).

Anhand der oben genannten statistischen Verfahren erfolgte die Auswertung und Berechnung, wobei alle abhängigen Variablen getrennt für den ersten und den zweiten Befragungszeitpunkt ausgezählt wurden und zusätzlich für alle Fragestellungen der Untersuchung zusätzlich zwischen den beiden Versuchsmaterialien (Dia-Serie 1 – 'PKW-PKW' und 2 –

‘PKW-Fahrrad’ bzw. Dia-Serie 1 – ‘PKW-PKW’ und ‘selbst erfundener Unfall-Serie’) unterschieden wurde.

Bei der Analyse aller Signifikanztests wurde von einem α – Niveau von 5 % ausgegangen.

3.1.7.2 Der Effektstärkenindex d beim Vergleich zweier Stichprobenmittelwerte

Die Effektstärke oder auch Effektgröße ist definiert als Größe eines Effekts bzw. einer Parameterdifferenz, in diesem Fall der Stichprobenmittelwerte. Es kann also damit angegeben werden, welcher Unterschied zwischen zwei Populationen (Gruppen) mindestens bestehen muss, um von einem praktisch bedeutsamen Unterschied sprechen zu können. Der Effektstärkenindex d gibt nun den Grad dieses Unterschieds an. In der Regel wird d für intervallskalierte Daten berechnet, die mittels parametrischer Verfahren miteinander verglichen werden.

d bestimmt sich wie folgt: $d = |\mu_1 - \mu_2| / \sigma$

wobei: μ_1 und μ_2 : Populationsmittelwerte

σ : Standardabweichung

(diese wird für beide Gruppen aufgrund der Varianzhomogenität als gleich angenommen)

3.2 Untersuchung zur Glaubhaftigkeit von Zeugenaussagen

3.2.1 Versuchspersonen

Als Versuchspersonen nahmen an dieser zweiten Teiluntersuchung der vorliegenden Dissertation 25 Jugendliche mit einer Intelligenzminderung im Alter zwischen ebenfalls 16 und 19 teil (MW = 18,35; SD = 0,80 zum ersten Befragungszeitpunkt T1).

Diese ebenfalls im B.B.W. Abensberg lebenden Jugendlichen hatten als Versuchspersonen alle bereits an der ersten Untersuchung teilgenommen, wurden somit aus dieser Stichprobe ausgewählt (siehe Kapitel 3.1.1). Dabei wurde bei der Zusammensetzung der Versuchsgruppe bzw. Auswahl der Versuchspersonen darauf geachtet, ein in etwa ausgewogenes Verhältnis der Geschlechter [14 männlich (56 %) - 11 weiblich (44 %)] sowie eine in Hinblick auf die bis dahin bekannten Ergebnisse und Eindrücke der ersten Teiluntersuchung repräsentative Zusammensetzung bezüglich des Intelligenzniveaus bzw. der kognitiven Kompetenz zu erhalten.

Abbildung 3-2: Verteilung der IQ-Werte (Glaubhaftigkeit)

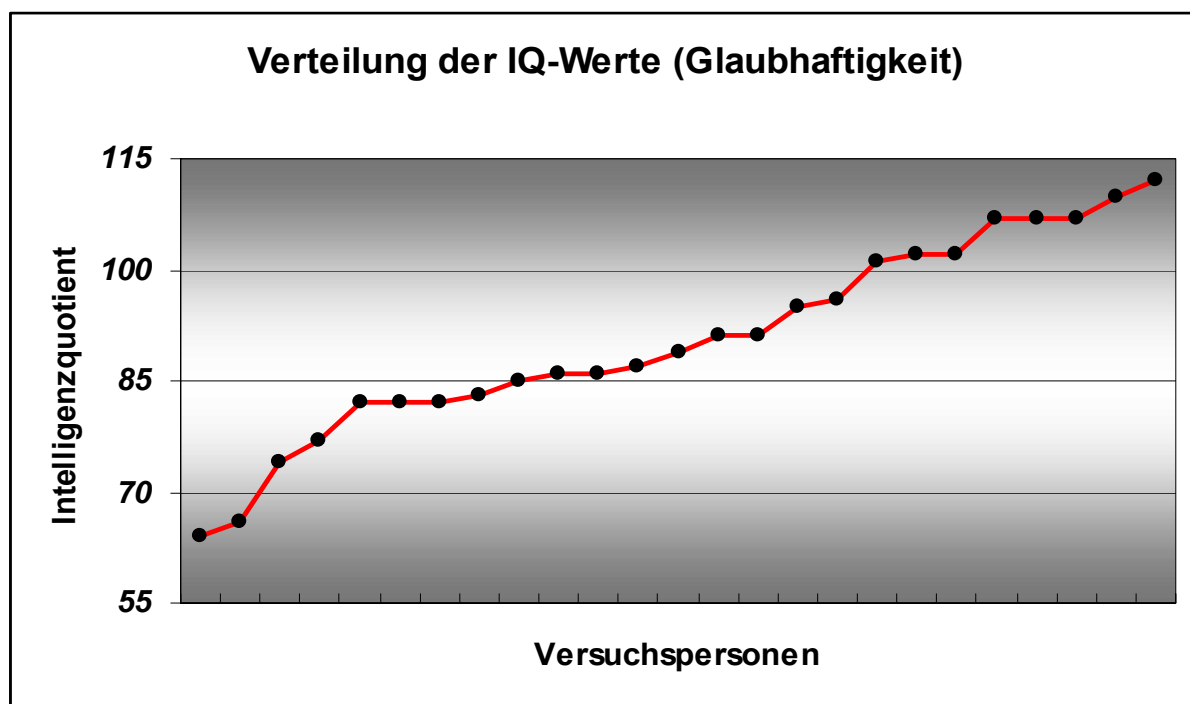


Tabelle 3-5: IQ-Daten 2 (Untersuchung 'Glaubhaftigkeit')

N	Mittelwert	SD	Median	Minimum	Maximum	Spannweite
25	90,56	13,23	89,00	64	112	48

Somit kann auch hier von einem relativ hohen Intelligenzniveau ausgegangen werden.

3.2.2 Grundlage für die Befragung und Instruktion der Versuchspersonen

3.2.2.1 Konzeption als Simulationsstudie

Simulationsstudien zeichnen sich dadurch aus, dass die Versuchspersonen darin Schilderungen oder Berichte über erlebnis-fundierte (wahre) aber auch erfundene, nicht-erlebnis-fundierte Geschehnisse und Vorfälle mit einer anderen, jedoch inhaltlichen ähnlichen Thematik äußern, also ein dynamischer Prozess in einem realen Modell nachgebildet wird. Von den Ergebnissen aus der Simulationsstudie wird anschließend wieder auf das Verhalten der realen Vorgänge zurück geschlossen. Eine von Steller, Wellershaus & Wolf 1992 durchgeführte Simulationsstudie zur Validität der Undeutsch-Hypothese ergab als eine wichtige Erkenntnis, dass

ein im Hinblick auf die externe Validität der wissenschaftlichen Untersuchung für den forensischen Bereich (...) ein Paradigma zu finden [war], das es erlaubt, Versuchspersonen zu Schilderung von Sachverhalten aufzufordern, für die nachweislich eine Erlebnisgrundlage gegeben ist oder nicht, wobei die emotionale und kognitive Bedeutung dieser Sachverhalte für den Aussagenden möglichst vergleichbar sein sollte mit dem Erleben eines Sexualdelikts, ohne dabei die Grenzen ethischer Zumutbarkeit für die Versuchspersonen zu überschreiten.

(STELLER, WELLERSHAUS & WOLF, 1992, S. 161).

3.2.2.2 Grundvariablen (Kriterien) der und Themen für die Geschichten

Demnach sollten im Rahmen von Simulationsstudien bei der Themenwahl nach STELLER, WELLERSHAUS & WOLF (1992) folgende Aspekte besonders berücksichtigt werden:

- Eigenbeteiligung des Aussagenden am Geschehen,
- eine vorwiegend negative emotionale Tönung des Geschehens sowie
- der weitgehende Kontrollverlust über die Situation

Dies wurde unter anderem von RAPP (1999) in Analogie aufgegriffen und entsprechend formuliert. Dabei sind folgende Gesichtspunkte von Bedeutung:

- das Ausmaß an Involviertheit der Versuchsperson (z.B. beobachtend vs. aktiv teilnehmend),
- der Grad an Vertrautheit mit dem Ereignis,
- die Komplexität des Ereignisses sowie
- die Nähe des Ereignisses zu einer realen Situation

3.2.2.3 Wahl der Themen und Vorgehensweise

STELLER et al. beschreiben dabei sieben Themen bzw. Ereignisse, die diese Grundvariablen beinhalten: „1) eine Injektion bekommen; 2) eine Operation als Patient erfahren; 3) Blut abgenommen bekommen; 4) Zähne gezogen oder Löcher gebohrt bekommen; 5) von einem Tier angefallen werden; 6) von einem anderen Kind verhauen werden; 7) einen Unfall erleiden, der eine medizinische Behandlung erfordert“ (STELLER, WELLERSHAUS & WOLF, 1992, S. 161).

Aus Gründen der Verständlichkeit wurden die in dieser Studie für die spezielle Gruppe der Versuchspersonen verwendeten Themen teilweise adäquat sprachlich und inhaltlich konkretisiert bzw. den Jugendlichen verständlich gemacht. Bei der inhaltlichen Präzisierung wurde darauf geachtet, dass die drei Grundvariablen weiterhin erfüllt sind (siehe *Anhang E bzw. F*).

Tabelle 3-6 stellt die Themen von STELLER et al. denen in dieser Studie verwendeten in erkennbarer Analogie gegenüber.

Tabelle 3-6: Gegenüberstellung der Themenkataloge von STELLER et al. (1992) vs. des hier verwendeten Katalogs

Themen nach STELLER et al. (1992)	Themen in dieser Studie
Injektion bekommen	Arztbesuch mit Spritze oder Blut abnehmen
Blut abgenommen bekommen	
Operation als Patient erfahren	Operation im Krankenhaus
Zähne gezogen oder Löcher gebohrt bekommen	Zahnarztbesuch mit Bohren oder Zahn ziehen
Von einem Tier angefallen werden	von einem Hund/Tier gebissen und verletzt werden
Von einem anderen Kind verhauen werden	Prügelei mit Gleichaltrigen
Einen Unfall erleiden, der eine medizinische Behandlung erfordert	Sturz aus größerer Höhe erleiden
	Arm oder Bein brechen/ verstauchen
	Verkehrsunfall erleiden
	sich mit Messer oder Schere schneiden
	barfuss in etwas (Biene, Scherbe) hineintreten und sich dabei verletzen

Die Instruktion der Versuchspersonen erfolgte ca. eine Woche vor dem ersten Befragungszeitpunkt T1. Dabei führte die Betreuerin dieser Arbeit, Frau Dr. Sandra Loohs, mit den Jugendlichen Einzelgespräche, um die Themen für die erfundenen und für die erlebnisbasierten Geschichten auszuwählen – natürlich erfolgte dies im Sinne der diesbezüglichen Blindheit des Verfassers dieser Arbeit ohne dessen Anwesenheit. Die Instruktion der Versuchspersonen erfolgte in Einzelgesprächen, um Verständnisproblemen und Kommunikationsschwierigkeiten vorzubeugen. Zusammen mit dem Jugendlichen wurden dann von Frau Dr. Loohs jeweils eine erlebnis-basierte sowie eine erfundene Geschichte aus dem Themenkreis festgelegt; Grundlage hierfür war ein Protokollbogen (siehe *Anhang E bzw. F*). Dabei sollte jede(r) Jugendliche anhand der oben genannten Kriterien und Themenauswahl eine Woche später dem Interviewer (Verfasser der Arbeit) sowohl eine selbst erlebte als auch eine erfundene Geschichte so glaubhaft und plausibel wie möglich schildern. Dabei wurden auch individuell Tipps und Anhaltspunkte gegeben, die dem Interviewer die Unterscheidung erschweren sollten.

3.2.3 Experimentalplan

Ziele dieser Teiluntersuchung waren, zu validieren, ob und wie gut die Aussagemerkmale (Realkennzeichen) zwischen erlebnisbasierten und nicht-erlebnisbasierten Aussagen von Jugendlichen mit einer Intelligenzminderung trennen können bzw. ob und welche Veränderungen in den Aussagen von einem Befragungszeitpunkt zum anderen festgestellt werden. Grundlage hierfür waren die transkribierten und 'blind gerateten' Aussagen.

3.2.3.1 Blindversuch

Um einen eventuellen Versuchsleitereffekt zu kontrollieren, war der Interviewer und Rater (Verfasser der Arbeit) zu beiden Befragungszeitpunkten sowie bei Auswertung der Befragungen blind bezüglich des Wahrheitsstatus der jeweiligen Aussagen/Geschichten. Er wusste also bis nach dem Ende aller Auswertungen zu keinem Zeitpunkt, welche der beiden Geschichten erlebnis-fundiert war und welche nicht.

3.2.3.2 Untersuchungsdesign

Die vorliegende Arbeit beinhaltet zudem ein längsschnittliches Design, d.h. die Versuchspersonen wurden im statistischen Sinn einer Messwiederholung aber auch eines praxisnahen forensischen Vorgehens zu zwei verschiedenen Befragungszeitpunkten, die ca. sechs Wochen auseinander lagen, interviewt.

Als unabhängige Variablen wurden zum einen der Faktor 'Befragungszeitpunkt', zum anderen der 'Wahrheitsstatus' oder die 'Glaubhaftigkeit' der Aussagen systematisch variiert.

Abhängige Variablen waren die letztendlich in die statistische Auswertung einbezogenen 17 Realkennzeichen (Aussagemerkmale).

Da jede der 25 Versuchspersonen zu beiden Befragungszeitpunkten je zwei Geschichten (eine erlebnisfundierte und eine nicht-erlebnisfundierte) erzählte, ergibt sich somit ein 2x2-faktorielles 'within-subject' - Design mit folgendem Experimentalplan:

Tabelle 3-7: Experimentalplan (2) - 'Glaubhaftigkeit'

Geschichte bzw. Wahrheitsstatus	Befragungszeitpunkt		Anzahl der Aussagen
	T1	T2	
<i>erlebnisfundierte</i>	25	25	50
<i>nicht erlebnisfundierte</i>	25	25	50
Anzahl der Aussagen	50	50	100

3.2.4 Durchführung der Befragungen

3.2.4.1 Befragungszeitpunkte

Eine Woche nach Instruktion der Jugendlichen durch Frau Dr. Loohs wurde mit der Befragung zum ersten Zeitpunkt T1 (Anfang Dezember 2003) begonnen. Der zweite Befragungszeitpunkt T2 fand ca. sechs Wochen nach T1 (Mitte Januar 2004) statt. Die Durchführung der jeweils 25 Interviews beanspruchte pro Befragungszeitraum ca. sechs Tage. Dabei wurde wiederum darauf geachtet, dass die Reihenfolge der Befragungen tageweise eingehalten wurde, also jede(r) Jugendliche denselben Delay zwischen Instruktion und jeweiligem Befragungszeitpunkt hatte.

3.2.4.2 Ablauf der Befragungen

Organisation, Ablauf, Struktur und Dauer der Interviews waren zu beiden Befragungszeitpunkten jeweils in etwa identisch.

Wie bei Teiluntersuchung 1 auch, wurden die Jugendlichen nacheinander zu einem Konferenzraum der Einrichtung bestellt und dort einzeln befragt. Dabei saßen sich Interviewer und Versuchsperson an einem Tisch gegenüber, die Befragung wurde mittels Diktiergerät und Konferenzmikrofon auf Tonband aufgenommen und stichpunktartig mitnotiert. Die Interviewdauer betrug durchschnittlich ca. 1 Stunde.

Nach kurzer Aufklärung über den Ablauf und Zweck der Befragung und der Möglichkeit der Klärung eventueller Unklarheiten begann die eigentliche Interviewphase. Die Versuchsperson sollte nun die beiden Geschichten nacheinander erzählen und konnte dabei frei entscheiden, welche der beiden Geschichten sie zuerst erzählen wollte.

Als Interviewform wurde gemäß den Hinweisen von KÖHNKEN et al. (1991) bzw. KÖHNKEN (1995) wiederum ein halbstrukturiertes Interview in Anlehnung einiger Elemente aus dem Kognitiven Interview von 'GEISELMAN & FISHER' gewählt, allerdings ohne die Verwendung der CI-spezifischen Gedächtnishilfen. Hauptelemente der Befragung waren daher folgende:

- *Begrüßung und Herstellen von Rapport*
- *Free Recall:*
Die Versuchsperson wurde zunächst aufgefordert, (im freien Bericht) alles zu berichten, was sie noch zu den Erlebnissen oder Vorfällen wusste.
- *Konkretes Nachfragen*
Hier wurde versucht, durch offen formulierte Fragen nähere Informationen zum Kontext, zu Interaktionen, zu Gesprächen, Gefühlen und zu weiteren Details zu erhalten und sich ein klares Bild zu verschaffen. Frageformen wie 'Ja-Nein' - Fragen und Suggestivfragen wurden dabei vermieden.
- *Verabschiedung und Hinweis*
Nach Beendigung des eigentlichen Interviews erfolgte der Dank an die Versuchsperson für's erste Mitmachen mit dem Hinweis, dass in sechs Wochen eine zweite, gleichartige Befragung folge.

Einen ungefähren Einblick gibt auch ein in *Anhang B* befindlicher Interviewleitfaden, der so oder ähnlich als Anhaltspunkt für solche Befragungen gelten kann.

3.2.5 Auswertung der Befragungen

3.2.5.1 Transkription und Rating der Interviews

Alle 50 (25 x 2) Interviews wurden nach Abschluss der Befragungen vom Verfasser der Arbeit wortwörtlich transkribiert. Da bei den Befragungen keine technischen oder sonstigen Probleme oder Ausfälle auftraten, konnten alle Interviews bzw. deren Transkripte in die Auswertung einfließen. Somit lagen von den 25 Versuchspersonen jeweils zwei Aussagen zu zwei Zeitpunkten vor, insgesamt also 100 Aussagen.

Auf ein Ratertraining wurde verzichtet, da von der entsprechenden Kompetenz und Erfahrung des Verfassers (Diplomarbeit und gutachterliche Tätigkeit) diesbezüglich ausgegangen wurde.

Diese Aussagen bzw. die Aussagemerkmale (Realkennzeichen) wurden ebenfalls vom Verfasser der vorliegenden Arbeit mittels des in *Anhang H* befindlichen Ratingbogens anhand einer 5-stufigen Skala ausgewertet und geratet.

Die Einstufung auf der Skala erfolgte nach jeweiliger Ausprägung bzgl. Häufigkeit, Intensität und Relativierung an der Gesamtaussage. Die Stufen der Skala reichten von 0 bis 4, wobei folgende inhaltliche Bedeutung gegeben war:

0 = „nicht vorhanden“	1 = „schwach vorhanden“	2 = „mittel vorhanden“
3 = „stark vorhanden“	4 = „sehr stark vorhanden“	

Im Anschluss an das eigentliche Rating der einzelnen Realkennzeichen wurden noch zwei weitere Einschätzungen abgegeben. Zum einen bezog sich dies auf einer 9-stufigen bipolaren Skala auf die Einschätzung der Glaubhaftigkeit der Aussage insgesamt (1 = „unglaubwürdig“ bzw. 9 = „glaubwürdig“), zum anderen auf einer 6-stufigen Skala darauf, wie sicher die Urteilsfindung eingestuft wird (1 = „sehr unsicher“ bzw. 6 = „sehr sicher“).

3.2.5.2 Zur Problematik der Auswertung, des Ratingverfahrens und Skalenniveaus verbaler Berichte

In diesem Bereich der experimentellen Gedächtnispsychologie und der Psychologie der Zeugenaussage ergibt sich oft die Frage nach der Art und Weise der Erfassung, Quantifizierung und Operationalisierung der Berichte und Schilderungen der Versuchspersonen. Die Methoden des ‘Scoring’ oder ‘Rating’ haben sich unter Berücksichtigung der bekannten und diskutierten Kriterien und dabei zu beachtenden Aspekte jedoch bereits in vielen Validierungsstudien als sehr gut geeignetes Instrument zur Erfassung der inhaltlichen Qualität der Realkennzeichen erwiesen. Auf eine ausführliche Diskussion soll hier deshalb verzichtet werden. Dennoch kann hier festgestellt werden, dass wohl in jedem Fall von konzeptorien-

tiertem Rating denn von intuitivem Rating ausgegangen werden kann. Ein präzise festgelegtes Merkmal, das aufgrund theoretischer Überlegungen und/oder empirischer Voruntersuchungen für wichtig und zielgemäß erachtet wird, ist hier Gegenstand der Messung. Es wird also nicht dem Rater überlassen, was er unter einem Merkmal versteht. Es liegt ein Konzept vor und die vom Rater vorgenommene Einschätzung von Realitätsausschnitten soll an diesem Konzept orientiert eingeschätzt werden. Die gewonnenen Eindrücke der Rater beim konzeptorientierten Rating sind an der vorgegebenen Merkmalsdefinition des Forschers orientiert. (...) Die Merkmalsdefinition durch den Rater erfolgt aufgrund expliziter Theorien des Forschers über die Merkmale.

Die zur Beurteilung von Zeugenaussagen Verwendung findenden Ratingskalen, sind nach dieser Definition dem konzeptorientierten Rating zuzuordnen. Die Merkmalsdefinition erfolgt theorie- und zielgemäß und damit losgelöst von alltäglichen Wahrnehmungsgewohnheiten intuitiver Rater (vgl. PETERSEN, 1997).

Ebenso kann und soll in dieser Arbeit nicht ausführlich und klärend auf das auch bereits oft diskutierte messtheoretische Problem von Ratingskalen und deren Skalenniveau eingegangen werden. Hier sei auch nur kurz darauf verwiesen. Die genannten Aspekte hängen natürlich unmittelbar mit der Entscheidung zusammen, welches statistische Verfahren letztendlich bei der Verwendung von Ratingskalen gerechnet werden darf, soll oder muss. Bei der Begründung der Angemessenheit von statistischen Verfahren muss zwischen messtheoretischen Interpretationsproblemen und mathematisch-statistischen Voraussetzungen unterschieden werden. Bei Ratingskalen ist zu bedenken, dass die Frage, ob verschiedene Zahlen unterschiedliche Ausprägungen des untersuchten Merkmals abbilden bzw. gleiche Zahlendifferenzen auch gleiche Merkmalsunterschiede repräsentieren, eher als messtheoretisches Problem und nicht als statistisches Problem zu sehen ist (vgl. BORTZ & DÖRING, 2003).

Nach BORTZ und DÖRING (2003) hätten parametrische Verfahren wie der t-Test oder die Varianzanalyse nicht intervallskalierte Daten zur Voraussetzung, sondern andere mathematische Voraussetzungen (wie normalverteilte, unabhängige und homogene Fehlerkomponenten). Diese parametrischen Verfahren sagen nichts über die Skaleneigenschaften aus.

Im Rahmen der Diskussion der statistischen „Puristen“ vs. die „Pragmatiker“ und um die Gratwanderung zwischen einer gewissen Entscheidungspragmatik und der eventuellen Verletzung der geforderten mathematisch-statistischen Voraussetzungen (vgl. auch BORTZ, LIENERT und BOEHNKE, 1990) zu verdeutlichen, wird auch in der vorliegenden Arbeit der pragmatische Weg vertreten und eingeschlagen und zum Teil auch parametrische Verfahren (ALM, t-Test) verwendet. Verteilungs- und andere Voraussetzungen wurden jeweils überprüft und daraufhin entschieden, welches Verfahren zu rechnen war. Die entsprechenden Ergebnisse sollen und werden hier dementsprechend vorsichtig und mit eventuellen Einschränkungen betrachtet und interpretiert.

3.2.5.3 Das Prinzip der Aggregation

Grundlage und Ausgangspunkt dieses auch in der vorliegenden Arbeit verwendeten Prinzips sind die diesbezüglichen Ausführungen von FIEDLER & SCHMID (1999), die im Rahmen ihres „Gutachtens über Methodik und Bewertungskriterien für Psychologische Glaubwürdigkeitsgutachten“ den induktiv-statistischen Schluss als eine Möglichkeit der wissenschaftlichen Begründung hierfür beschreiben. Bei dieser Methode „werden viele schwach bestätigte Indikatoren auf unterer Ebene benutzt“. Obwohl die Validität der einzelnen Indikatoren „bescheiden“ sein kann, kann „die aus der Gesamtheit aller Indikatoren abgeleitete Schlussfolgerung eines Gutachtens jedoch einen beträchtlichen diagnostischen Wert erreichen, der weit höher liegt als die Gültigkeit der einzelnen schwachen Indikatoren“ (FIEDLER & SCHMID, 1999, S. 11). Aus diesem Grund ist es zulässig, verschiedene Realkennzeichen (Indikatoren) sowohl nach Kategorien als auch als Ganzes zusammenzufassen. Nach DAHLE (1997) muss hier allerdings kritisch angemerkt werden, dass die Realkennzeichen dann jedoch „ungewichtet in die Summenbildung“ eingehen (vgl. DAHLE, 1997, S. 21).

Somit wurden auch in dieser Arbeit sowohl Summenscores in Form von eigenen abhängigen Variablen für die einzelnen Merkmalskategorien (Allgemeine Merkmale, Spezielle Inhalte, Inhaltliche Besonderheiten und Motivationsbezogene Inhalte) gesamt als auch zu den verschiedenen Befragungszeitpunkten gebildet. Anhand des ALM (Allgemeines Lineares Modell) sowie der post-hoc durchgeführten T-Test und des eben erläuterten Aggregationsprinzip lassen sich nun Aussagen darüber treffen, ob sich gegebenenfalls ganze Merkmalskategorien bzw. alle Realkennzeichen als Gesamtinstrument zur Diskrimination zwischen erlebnisbasierten und nicht-erlebnisbasierten Aussagen eignen oder einen gewissen Indikatorwert aufweisen.

3.2.6 Variablenübersicht

Wie bereits erwähnt, wurden im Experiment zwei unabhängige Variablen systematisch variiert: der Wahrheitsstatus der Aussagen und der Befragungszeitpunkt.

Zudem finden sich folgende Moderatorvariablen: Delay der Befragung, Berufs- bzw. Ausbildungszweig, Geschlecht der Versuchspersonen.

Nach Auswertung aller Interviews ergaben sich zudem letztlich insgesamt 23 bzw. 138 (23 x 6 Abstufungen) abhängige Variablen (die Realkennzeichen im einzelnen, in den Merkmalskategorien sowie als Ganzes), bei denen jeweils von einem intervallskalierten Niveau ausgegangen wird:

Dabei bedeuten:

- .wT1/2 = Ausprägung der Variable der wahren Geschichte zu T1/2
- .eT1/2 = Ausprägung der Variable der erfundenen Geschichte zu T1/2
- .w = Ausprägung der Variable der wahren Geschichte über beide Befragungszeitpunkte
- .e = Ausprägung der Variable der wahren Geschichte über beide Befragungszeitpunkte

Tabelle 3-8: Abhängige Variablen (2) - 'Glaubhaftigkeitsuntersuchung'

VARIABLE bzw. KATEGORIE	VARIABLENFORM / - ABSTUFUNG					
einzelne Realkennzeichen (Aussagemerkmale)						
Allgemeine Merkmale						
Logische Konsistenz	.wT1	.eT1	.wT2	.eT2	.w	.e
Widerspruchsfreie, ungeordnete Darstellung	.wT1	.eT1	.wT2	.eT2	.w	.e
Quantitativer Detailreichtum	.wT1	.eT1	.wT2	.eT2	.w	.e
Spezielle Inhalte						
Kontextuelle Einbettung	.wT1	.eT1	.wT2	.eT2	.w	.e
Schilderung nonverbaler Aktionen	.wT1	.eT1	.wT2	.eT2	.w	.e
Wiedergabe von Gesprächen	.wT1	.eT1	.wT2	.eT2	.w	.e
Schilderung von Komplikationen im Handlungsverlauf	.wT1	.eT1	.wT2	.eT2	.w	.e
Inhaltliche Besonderheiten						
Nebensächliche Details	.wT1	.eT1	.wT2	.eT2	.w	.e
Ausgefallene Details	.wT1	.eT1	.wT2	.eT2	.w	.e
Schilderung eigener psychischer / physischer Vorgänge	.wT1	.eT1	.wT2	.eT2	.w	.e
Schilderung psychischer / physischer Vorgänge beim Täter oder Beteiligten	.wT1	.eT1	.wT2	.eT2	.w	.e
Motivationsbezogene Inhalte						
Spontane Verbesserungen der eigenen Aussage	.wT1	.eT1	.wT2	.eT2	.w	.e
Zugeben von Erinnerungslücken	.wT1	.eT1	.wT2	.eT2	.w	.e
Einwände gegen die Glaubhaftigkeit der eigenen Aussage	.wT1	.eT1	.wT2	.eT2	.w	.e
Selbstbelastungen	.wT1	.eT1	.wT2	.eT2	.w	.e
Inschutznahme des Täters oder Beteiligten	.wT1	.eT1	.wT2	.eT2	.w	.e
Deliktspezifische Inhalte						
Deliktspezifische Details oder Elemente	.wT1	.eT1	.wT2	.eT2	.w	.e
Merkmalskategorien						
Allgemeine Merkmale	.wT1	.eT1	.wT2	.eT2	.w	.e
Spezielle Inhalte	.wT1	.eT1	.wT2	.eT2	.w	.e
Inhaltliche Besonderheiten	.wT1	.eT1	.wT2	.eT2	.w	.e
Motivationsbezogene Inhalte	.wT1	.eT1	.wT2	.eT2	.w	.e
Deliktspezifische Inhalte	.wT1	.eT1	.wT2	.eT2	.w	.e
alle Realkennzeichen (Aussagemerkmale) gesamt						
RKZ total	.wT1	.eT1	.wT2	.eT2	.w	.e

3.2.7 Statistische Auswertung

Wiederum wurden die dazugehörigen statistischen Analysen anhand des Statistikprogramms 'SPSS für Windows - Version 12.0' durchgeführt; Abbildungen, Diagramme und graphische Darstellungen mittels Microsoft Office Excel 2003.

3.2.7.1 Signifikanztests

Analog zu den diesbezüglichen Ausführungen der 'source-monitoring' - Untersuchung wird bei Annahme intervallskalierter Daten zunächst die Verteilungsform für jede der Variablen bzw. deren Abstufungen (wT1, eT1, wT2 und eT2) anhand des Einstichproben-KOLMOGOROV-SMIRNOV-Test überprüft.

Als statistische Verfahren wurden folgende verwendet:

- a) Allgemeines Lineares Modell ('ALM')
- b) T – Tests für unabhängige Stichproben (post-hoc)
- c) Wilcoxon – Test für Paardifferenzen (post-hoc)

Der T-Test für unabhängige Stichproben erweist sich insgesamt dabei als relativ robust gegenüber eventuellen Verletzungen der geforderten Voraussetzungen (vgl. BORTZ, 1999).

Bei der Analyse aller Signifikanztests wurde von einem α – Niveau von 5 % ausgegangen.

3.2.7.2 Der Effektstärkenindex d beim Vergleich zweier Stichprobenmittelwerte

Für jedes einzelne Realkennzeichen sowie auch für die fünf Merkmalskategorien und alle Realkennzeichen insgesamt wurde die entsprechende Effektstärke berechnet.

Auf die diesbezüglichen theoretischen Ausführungen sei an dieser Stelle auf das analoge Kapitel der 'source-monitoring' - Untersuchung verwiesen.

3.2.7.3 Übersicht über die angewandten statistischen Verfahren

Tabelle 3-9: angewandte statistische Verfahren (einzelne Realkennzeichen)

REALKENNZEICHEN (Aussagemerkmale)	Statistisches Verfahren	post - hoc Verfahren
Logische Konsistenz	ALM	W
Widerspruchsfreie, ungeordnete Darstellung	ALM	T, W
Quantitativer Detailreichtum	ALM	T
Kontextuelle Einbettung	ALM	T, W
Schilderung nonverbaler Aktionen	ALM	W
Wiedergabe von Gesprächen	ALM	T
Schilderung von Komplikationen im Handlungsverlauf	ALM	W
Ausgefallene Details	ALM	W
Nebensächliche Details	ALM	W
Beschreibung eigener psychischer oder physischer Vorgänge	ALM	T, W
Beschreibung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	ALM	W
Spontane Verbesserungen der eigenen Aussage	ALM	W
Zugeben von Erinnerungslücken	ALM	W
Einwände gegen die Richtigkeit der eigenen Aussage	ALM	W
Selbstbelastungen	ALM	W
Inschutznahme des Täters oder Beteiligten	ALM	W
Deliktsspezifische Details oder Elemente	ALM	W

Tabelle 3-10: angewandte statistische Verfahren (Merkmalskategorien)

MERKMALSKATEGORIE	Statistisches Verfahren	post – hoc Verfahren
Allgemeine Merkmale	ALM	T
Spezielle Inhalte	ALM	T
Inhaltliche Besonderheiten	ALM	T
Motivationsbezogene Inhalte	ALM	T
Deliktsspezifische Inhalte	ALM	T
Alle Realkennzeichen	ALM	T

wobei: T = T-Test für unabhängige Stichproben; W = Wilcoxon – Test für Paardifferenzen

4. Ergebnisse

4.1 Ergebnisse zur Erinnerungsleistung und dem source- bzw. reality monitoring

Im Folgenden werden – nachdem kurz auf Verteilungsform und Varianzhomogenität der Daten eingegangen wurde – die Ergebnisse der ersten Teiluntersuchung umfassend dargestellt.

Dabei wird zunächst auf die Unterschiede in der Zeugenkompetenz zwischen den beiden Versuchsgruppen zu den verschiedenen Befragungszeitpunkten eingegangen.

Im Anschluss daran folgt die Darstellung der Ergebnisse der Veränderungen der Gedächtnisleistungen über die Zeit, also die beiden Befragungszeitpunkte hinweg, getrennt für beide Versuchsgruppen.

Daraufhin werden die Ergebnisse zu den einzelnen Befragungszeitpunkten im Vergleich zueinander, jeweils wieder separat für beide Versuchsbedingungen erläutert.

Abschließend werden die Ergebnisse zu den Haupt- und Interaktionseffekten der Gruppen und der Befragungszeitpunkte dargestellt.

Zum begrifflichen Verständnis:

In der gesamten weiteren Darstellung werden folgende Begriffe jeweils gleichbedeutend zueinander verwendet:

- externe 'source monitoring' (ESM) -Gruppe ↔ external-externe Bedingung/Gruppe
- 'reality monitoring' (RM) - Gruppe ↔ external-interne Bedingung/Gruppe

4.1.1 Überprüfung der (intervallskalierten) Variablen auf Verteilungsform und Varianzhomogenität

4.1.1.1 Überprüfung der Verteilungsform

Die Variablen wurden jeweils getrennt nach den Versuchsgruppen und getrennt nach der Anzahl ihrer Befragungen zunächst mittels des Einstichproben – Kolmogorov-Smirnov - Tests auf ihre Normalverteilung hin überprüft. Dies dient dazu zu untersuchen, ob ein signifikanter Unterschied zwischen der vorliegenden empirischen Verteilung der Daten und der vorgegebenen theoretischen Verteilungsform festgestellt werden kann. Liegt ein nicht-signifikantes Testergebnis vor, kann davon ausgegangen werden, dass die Verteilung der Daten der theoretischen Verteilung hinreichend entspricht, also von einer Normalverteilung

der Daten ausgegangen werden kann. Im Folgenden sind die einzelnen Ergebnisse für die Variablen nach Gruppen und Anzahl der Befragungen angegeben:

(MD = maximale Differenz; K-S-Z = Prüfgröße des Kolmogorov-Smirnov – Tests; p = Irrtumswahrscheinlichkeit von 0,05 - bezogen auf eine zweiseitige asymptotische Signifikanz)

Tabelle 4-1: Ergebnisse des K-S – Tests auf Normalverteilung (1) für die Gruppe 'external-external' ('source-monitoring') für alle Variablen

Variable	T1			T2			NT2		
	MD	K-S-Z	p	MD	K-S-Z	p	MD	K-S-Z	p
korr.Details.gesamt	,179	0,593	0,874	,178	0,591	0,876	,154	0,510	0,957
korr.Details.freerecall	,191	0,633	0,818	,182	0,605	0,858	,114	0,378	0,999
korr.Details.Nachfrag	,178	0,589	0,878	,107	0,356	1,000	,122	0,404	0,997
Fehler gesamt	,210	0,695	0,720	,111	0,368	0,999	,131	0,434	0,992
QueVw.insgesamt	,202	0,668	0,763	,194	0,643	0,802	,251	0,833	0,491
QueVw.innerhalb	,382	1,267	0,081	,353	1,169	0,130	,395	1,309	0,065
QueVw.zwischen	,183	0,606	0,856	,248	0,821	0,510	,227	0,754	0,621
Schemat. Intrusionen	,318	1,055	0,215	,284	0,940	0,340	,300	0,994	0,276
Konfabulationen	,266	0,882	0,419	,259	0,860	0,451	,310	1,028	0,242
Allg.Erinnerungsfehl.	,176	0,584	0,885	,143	0,475	0,978	,172	0,572	0,899

Tabelle 4-2: Ergebnisse des K-S – Tests auf Normalverteilung (2) für die Gruppe 'external-internal' ('reality-monitoring') für alle Variablen

Variable	T1			T2			NT2		
	MD	K-S-Z	p	MD	K-S-Z	p	MD	K-S-Z	p
korr.Details.gesamt	,152	0,527	0,944	,144	0,500	0,964	,214	0,676	0,752
korr.Details.freerecall	,138	0,477	0,977	,150	0,519	0,951	,268	0,846	0,472
korr.Details.Nachfrag	,162	0,560	0,912	,230	0,796	0,551	,206	0,652	0,789
Fehler gesamt	,164	0,568	0,904	,137	0,473	0,979	,274	0,866	0,441
QueVw.insgesamt	,179	0,619	0,837	,268	0,930	0,353	,258	0,816	0,519
QueVw.innerhalb	,207	0,719	0,680	,310	1,075	0,198	,182	0,577	0,894
QueVw.zwischen	,401	1,390	0,042	,352	1,221	0,102	,346	1,093	0,183
schemat. Intrusionen	,209	0,726	0,668	,376	1,304	0,067	,260	0,823	0,507
Konfabulationen	,224	0,775	0,586	,162	0,562	0,910	,156	0,494	0,968
Allg.Erinnerungsfehl.	,191	0,661	0,775	,130	0,451	0,987	,218	0,689	0,730

Zusammenfassend kann somit festgestellt werden, dass bis auf eine Variable zu einem Befragungszeitpunkt der Gruppe 'reality monitoring' ('Quellenverwechslungen zwischen den Dia-Serien zu Befragungszeitpunkt T1') bei allen übrigen Variablen zu allen Befragungszeitpunkten und für die beiden Gruppen von einer Normalverteilungsannahme ausgegangen werden kann.

4.1.1.2 Überprüfung der Varianzhomogenität

Alle Variablen, für die wegen des oben genannten und durchgeführten K-S – Tests eine Normalverteilung angenommen werden kann, wurden des Weiteren auf ihre Varianzhomogenität hin überprüft. Dazu wurde der im Allgemeinen ohnehin als recht robust gegenüber Verletzungen der Normalverteilungsannahme geltende LEVENE-Test verwendet. Für den Fall eines nicht-signifikanten Ergebnisses des LEVENE-Test nimmt man an, dass sich die Fehlervarianzen der jeweiligen Versuchsgruppen nicht signifikant unterscheiden.

(F = LEVENE-Prüfgröße; df = Freiheitsgrade; p = Irrtumswahrscheinlichkeit von 0,05)

Tabelle 4-3: Varianzhomogenität (1) für alle Variablen zu Befragungszeitpunkt T1 und T2 über die Gruppen

Variable	T1				T2			
	F	df 1	df 2	p	F	df 1	df 2	p
korrekte Details gesamt	11,407	1	21	0,003	0,733	1	21	0,402
korr. Details im 'free recall'	1,209	1	21	0,284	0,062	1	21	0,806
korr. Details auf Nachfrage	5,861	1	21	0,025	0,129	1	21	0,723
Fehler gesamt	0,008	1	21	0,929	0,012	1	21	0,915
Quellenverwechslg.gesamt	0,405	1	21	0,531	1,614	1	21	0,218
QueVw.innerhalb.d.Serien	2,523	1	21	0,127	0,000	1	21	0,982
QueVw.zwischen.d.Serien	2,616	1	21	0,121	1,926	1	21	0,180
schematypisch.Intrusionen	0,802	1	21	0,381	0,002	1	21	0,969
Konfabulationen	26,016	1	21	< 0,001	2,111	1	21	0,161
Allgem. Erinnerungsfehler	0,896	1	21	0,355	0,223	1	21	0,642

Somit kann für fast alle Variablen die Varianzhomogenität angenommen werden, da der LEVENE-Test mit Ausnahme von drei Variablen für alle übrigen kein signifikantes Ergebnis ausweist. Die Anwendung varianzanalytischer bzw. metrischer Verfahren ist also zu rechtfertigen.

Für jeden t-Test wurde ebenfalls der LEVENE-Test zur Überprüfung der Varianzhomogenität durchgeführt. Im Fall der Gleichheit der Varianzen wurde der entsprechende T-Wert bzw. das entsprechende Signifikanz-Ergebnis verwendet, bei Nicht-Homogenität analog der andere Wert.

Alle beschriebenen Ergebnisse von Signifikanztests (t-Test) beziehen sich – soweit nicht explizit anders angegeben – auf eine *einseitige* Hypothesenprüfung. Die Ergebnisse des ALM (allgemeines lineares Modell) mit Messwiederholung (Ergebnisse zu den Haupt- und Interaktionseffekten) beziehen sich auf eine zweiseitige Hypothesenprüfung. Das ALM mit Messwiederholung entspricht in etwa einer univariaten Varianzanalyse mit Messwiederholung.

4.1.2 Ergebnisse der Unterschiede zwischen externaler ‘source monitoring’ - Gruppe (ESM-Gruppe) und ‘reality monitoring’ - Gruppe (RM-Gruppe)

Anhand eines t-Tests für unabhängige Stichproben wurde geprüft, ob und welche Unterschiede in der Gedächtnis- und Wiedergabeleistung sich zwischen den beiden Versuchsgruppen finden lassen.

Die Ergebnisse der T-Tests beziehen sich für die einzelnen Befragungszeitpunkte T1, T2 und NT2 (Kapitel 4.1.2) auf eine zweiseitige Hypothesenprüfung.

Dabei werden im Folgenden zu den jeweiligen Befragungszeitpunkten zunächst die Ergebnisse über beide Dia- bzw. Unfall-Serien hinweg dargestellt. Im Anschluss daran folgen jeweils zunächst die Ergebnisse für die Dia-Serie ‘PKW-PKW’, danach die Ergebnisse für die Serie ‘PKW-Fahrrad’.

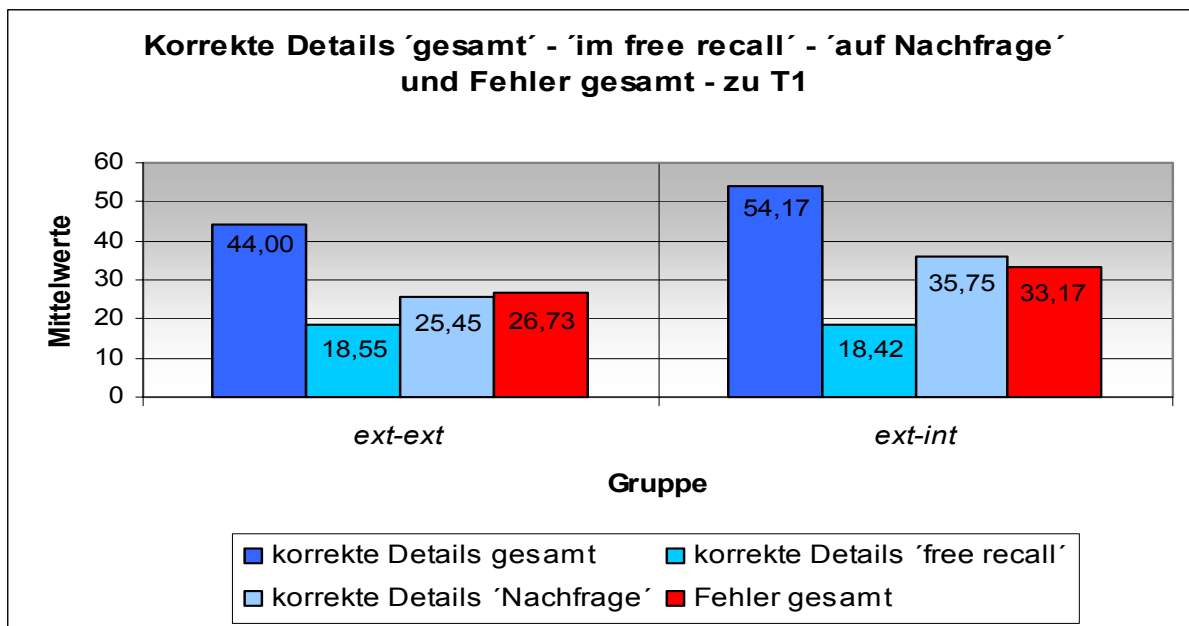
4.1.2.1 Ergebnisse zum ersten Befragungszeitpunkt T1

Zu T1 wurden insgesamt N = 23 Versuchspersonen in beiden Gruppen interviewt, wobei die externe ‘source monitoring’ – Gruppe 11 Versuchspersonen umfasst und die ‘reality monitoring’ – Gruppe 12 Versuchspersonen.

Die erste Befragung zu T1 fand für diese Versuchspersonen eine Woche nach Präsentation des Dia-Materials statt.

Betrachtet man dabei zunächst die berichteten korrekten Details sowie die insgesamt begangenen Fehler, ergaben sich dabei folgende Ergebnisse:

Abbildung 4-1: Vergleich der beiden Gruppen hinsichtlich der korrekten Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zum ersten Befragungszeitpunkt T1 – für beide Serien



Die Gruppe 'external-internal' berichtet dabei zwar mehr korrekte Details insgesamt als die Gruppe 'external-external', jedoch erweist sich dieser Unterschied mit ($T(15,888) = -1,809$; $p = 0,089$) als statistisch nicht signifikant. Im Ergebnis ebenso nicht signifikant ($T(21) = 0,035$; $p = 0,972$) verhält es sich für die jedoch fast identischen Mittelwerte bezüglich der korrekten Details im 'free recall'. Ein signifikanter Unterschied findet sich allerdings für die auf Nachfrage korrekt berichteten Details. Hier weist der t-Test ein signifikantes Ergebnis aus ($T(21) = 0,035$; $p = 0,017$).

Betrachtet man die hingegen die insgesamt begangenen Fehler, so lässt sich feststellen, dass die 'reality monitoring' - Gruppe zum ersten Befragungszeitpunkt deutlich mehr Fehler macht, als die externe 'source monitoring' - Gruppe. Dieser Unterschied erweist sich mit $T(20,487) = -2,595$; $p = 0,017$) als signifikant.

Tabelle 4-4: Mittelwerte und Standardabweichungen der beiden Gruppen für korrekte Details ('gesamt', 'free recall' und 'auf Nachfrage') und 'Fehler gesamt' zum ersten Befragungszeitpunkt T1 – für beide Serien

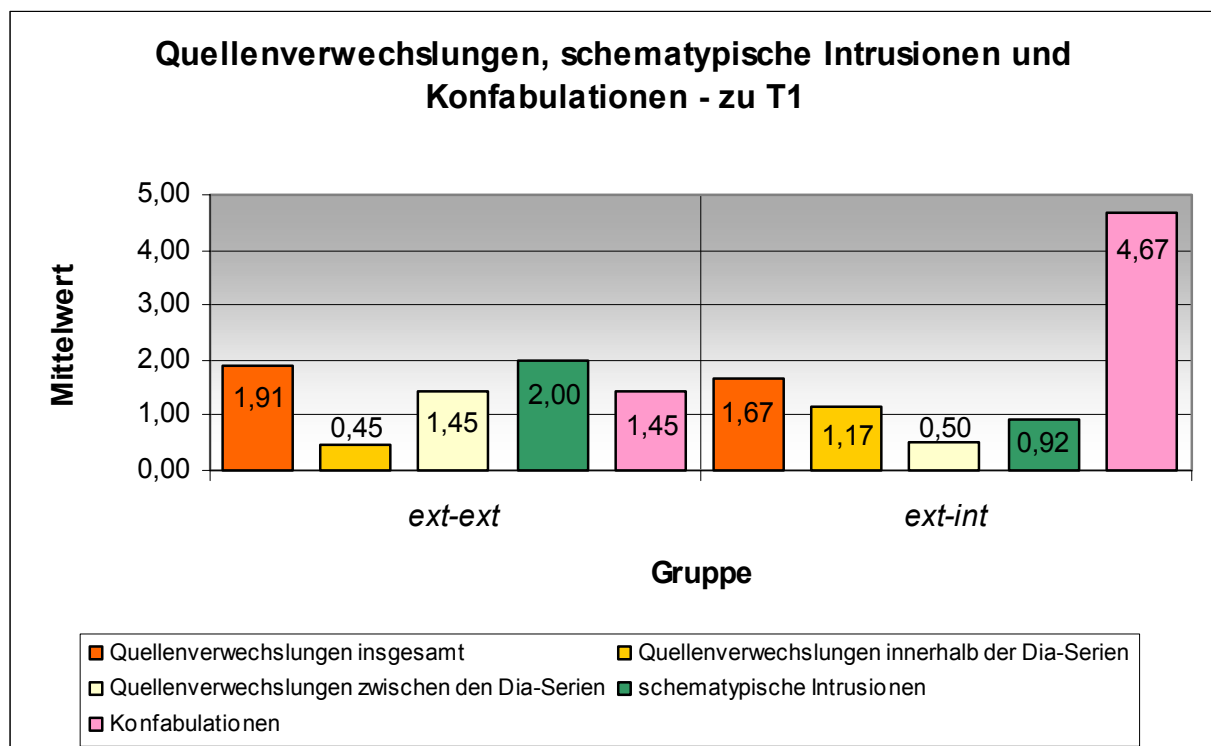
Gruppe	k. Det. 'gesamt'	k. Det. 'free recall'	k. Det. 'Nachfrage'	Fehler gesamt
ext-ext	MW = 44,00 SD = 8,19	MW = 18,55 SD = 7,56	MW = 25,45 SD = 5,99	MW = 26,73 SD = 6,13
ext-int	MW = 54,17 SD = 17,49	MW = 18,42 SD = 9,69	MW = 35,75 SD = 13,90	MW = 33,17 SD = 5,73

Tabelle 4-5: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zum ersten Befragungszeitpunkt T1 – für beide Serien

Variable	T	df	Signifikanz
korrekte Details gesamt	-1,809	15,888	0,089
korrekte Details im 'free recall'	0,035	21	0,972
korrekte Details 'auf Nachfrage'	-2,340	15,222	0,033
Fehler gesamt	-2,595	20,487	0,017

Bezüglich der Quellenverwechslungen, der schematypischen Intrusionen, der Konfabulationen sowie der allgemeinen Erinnerungsfehler ergibt sich folgendes Bild:

Abbildung 4-2: Vergleich der beiden Gruppen hinsichtlich der Quellenverwechslung insgesamt sowie beider Arten der Quellenwechslung, der schematypischen Intrusionen und der Konfabulationen zum ersten Befragungszeitpunkt T1 – für beide Serien



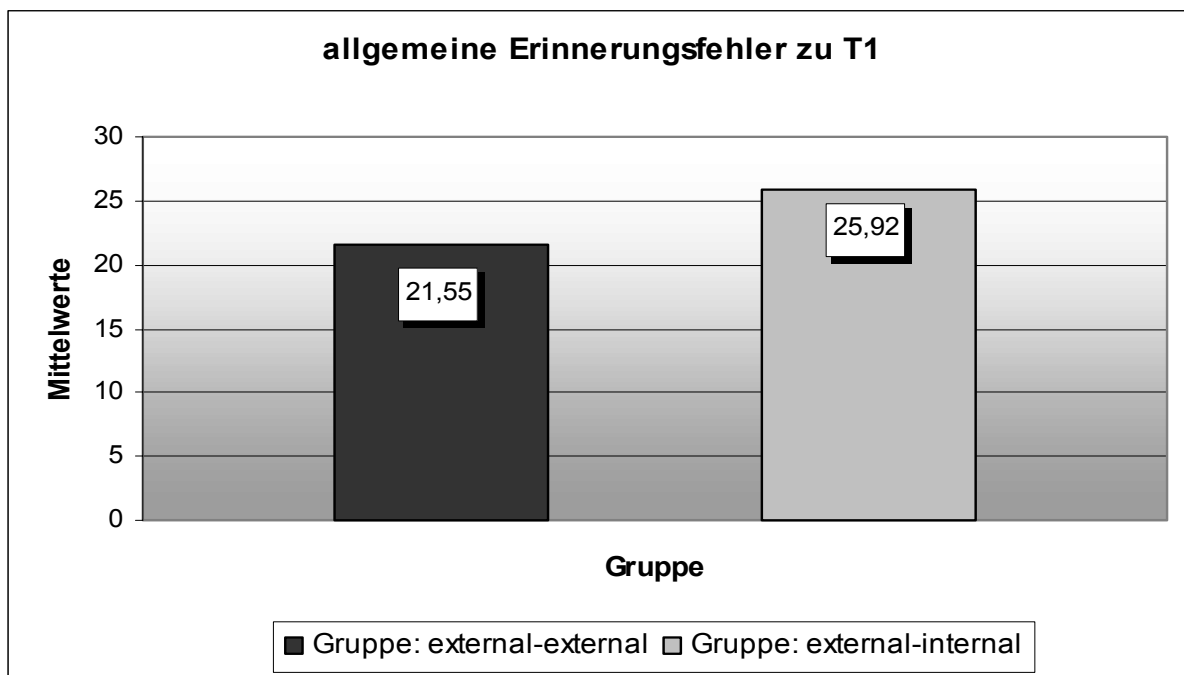
Was die insgesamt begangenen Quellenverwechslungen angeht, ist ersichtlich, dass die Gruppe 'reality monitoring' zwar tendenziell etwas weniger Fehler macht, dieser Unterschied aber nicht signifikant ist ($T(21) = 0,378$; $p = 0,709$). Bei differenzierter Betrachtung fällt hier jedoch auf, dass innerhalb der Dia-Serien von der Gruppe 'external-external' weniger Quel-

lenverwechslungen begangen werden als von der Gruppe 'external-internal', zwischen den Dia-Serien verhält es sich hingegen umgekehrt. Die Unterschiede innerhalb der Dia-Serien sind jedoch statistisch unbedeutsam ($T(21) = -1,931$; $p = 0,067$). Für die Quellenverwechslungen zwischen den Serien kann mit $T(21) = 2,151$; $p = 0,043$ allerdings eine signifikante Veränderung der Mittelwerte beobachtet werden.

Schematypische Intrusionen finden sich in der ESM-Gruppe signifikant mehr als in der RM-Gruppe ($T(21) = 2,292$; $p = 0,032$).

In der Bedingung 'external-internal' konfabulieren die Versuchspersonen zum ersten Befragungszeitpunkt mit im Durchschnitt über drei Fehlern mehr in einem wesentlich höheren Maß als in der Bedingung 'external-external', was statistisch zu einem hochsignifikanten Ergebnis führt ($T(13,601) = -3,744$; $p = 0,002$).

Abbildung 4-3: Vergleich der beiden Gruppen hinsichtlich der allgemeinen Erinnerungsfehler zum ersten Befragungszeitpunkt T1 – für beide Serien



Ein statistisch unbedeutsames Ergebnis lässt sich auch bezüglich der allgemeinen Erinnerungsfehler konstatieren. Hier erinnert die Gruppe 'external-internal' durchschnittlich etwas mehr als vier Aspekte und Details weniger als die Gruppe 'external-external'. Dies zeigt sich in einem nicht-signifikanten Ergebnis des t-Tests für unabhängige Stichproben ($T(21) = -1,743$ und $p = 0,096$).

Tabelle 4-6: Mittelwerte und Standardabweichungen der beiden Gruppen für diese Variablen zum ersten Befragungszeitpunkt T1 – für beide Serien

Gruppe	QV total	QV intra	QV inter	Intrusionen	Konfabulation	AEF
ext-ext	MW = 1,91 SD = 1,64	MW = 0,45 SD = 0,69	MW = 1,45 SD = 1,29	MW = 2,00 SD = 1,41	MW = 1,45 SD = 0,93	MW = 21,55 SD = 6,74
ext-int	MW = 1,67 SD = 1,44	MW = 1,17 SD = 1,03	MW = 0,50 SD = 0,80	MW = 0,92 SD = 0,79	MW = 4,67 SD = 2,81	MW = 25,92 SD = 5,25

Tabelle 4-7: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen Quellenverwechslungen innerhalb und zwischen den Dia-Serien, schematypische Intrusionen, Konfabulationen und allgemeine Erinnerungsfehler zum ersten Befragungszeitpunkt T1 – für beide Serien

Variable	T	df	Signifikanz
Quellenverwechslungen insgesamt	0,378	21	0,709
Quellenverwechslungen innerhalb	-1,931	21	0,067
Quellenverwechslungen zwischen	2,151	21	0,043
schematypische Intrusionen	2,292	15,429	0,032
Konfabulationen	-3,744	13,601	0,002
allgemeine Erinnerungsfehler	-1,743	21	0,096

Betrachtet man nun die beiden Unfall-Serien getrennt voneinander ergibt sich folgendes Bild:

Dia-Serie 1 (PKW-PKW):

Was das Erinnern bzw. Berichten korrekter Details insgesamt wie auch auf Nachfrage betrifft, lässt sich feststellen, dass zum ersten Befragungszeitpunkt die RM – Gruppe deutlich besser ist als die ESM - Gruppe. Beides zeigt sich in höchst- bzw. hochsignifikanten Ergebnissen ($T(21) = -3,078$ und $p = 0,001$ bzw. $T(21) = -3,165$ und $p = 0,005$). Bei den im freien Bericht erinnerten Details erweisen sich die Unterschiede allerdings als so gering, dass sich entsprechend kein signifikantes Ergebnis ergab.

Ebenso wenig bedeutsam scheinen die Unterschiede zwischen den beiden Versuchsgruppen bezüglich der Dia-Serie 'PKW-PKW' zu sein, was die Gesamtfehlerzahl sowie die Zahl der allgemeinen Erinnerungsfehler, der schematypischen Intrusionen sowie der berichteten Konfabulationen betrifft – zumindest noch zum ersten Befragungszeitpunkt. Hierfür konnten

jeweils keine signifikanten Mittelwertsunterschiede konstatiert werden. Tendenziell jedoch lagen bei diesen Fehlerarten mit Ausnahme der Intrusionen die Werte der RM – Gruppe höher.

Zwar fanden sich hinsichtlich der insgesamt und innerhalb der Dia-Serie begangenen Quellenverwechslungen keine statistisch bedeutsamen Unterschiede zwischen den Versuchsbedingungen, jedoch zu T1 mit einem fast hochsignifikanten Testergebnis ($T(10,000) = 3,105$ und $p = 0,011$) sehr wohl für die Quellenverwechslungen zwischen den beiden Unfall-Serien. Diejenigen Versuchspersonen, die sich zusätzlich die zweite Unfall-Serie selbst ausdenken mussten (Bedingung 'external-internal') konnten bei Abfrage der Dia-Serie 'PKW-PKW' diese - im Vergleich zu den Versuchspersonen der Bedingung 'external-external' - offensichtlich sehr gut bzw. besser von der selbst erfundenen Unfall-Serie differenzieren und begingen keinerlei Quellenverwechslungen. Die Versuchspersonen der Gruppe ESM begingen hier zwischen den beiden von ihnen gesehenen Dia-Serien mehr Quellenverwechslungen und konnten diese und deren Inhalt demnach nicht so gut differenzieren.

Die *Tabellen 4-8 und 4-9* verdeutlichen die eben erläuterten Ergebnisse im Detail (Mittelwerte, Standardabweichungen sowie Ergebnisse der t-Tests für unabhängige Stichproben):

Tabelle 4-8: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum ersten Befragungszeitpunkt T1 – für Serie 'PKW-PKW'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 23,18; SD = 6,27	MW = 35,83; SD = 9,58
korrekte Details im 'free recall'	MW = 8,82; SD = 3,87	MW = 11,17; SD = 5,10
korrekte Details 'auf Nachfrage'	MW = 14,36; SD = 6,99	MW = 24,67; SD = 8,47
Fehler gesamt	MW = 14,27; SD = 5,14	MW = 15,67; SD = 3,06
Quellenverwechslungen insgesamt	MW = 1,09; SD = 1,04	MW = 0,67; SD = 0,99
Quellenverwechslungen innerhalb	MW = 0,27; SD = 0,47	MW = 0,67; SD = 0,99
Quellenverwechslungen zwischen	MW = 0,82; SD = 0,87	MW = 0,00; SD = 0,00
schematypische Intrusionen	MW = 0,91; SD = 0,83	MW = 0,50; SD = 0,67
Konfabulationen	MW = 0,91; SD = 0,94	MW = 1,75; SD = 1,49
allgemeine Erinnerungsfehler	MW = 11,36; SD = 5,28	MW = 12,75; SD = 3,19

Tabelle 4-9: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum ersten Befragungszeitpunkt T1 – für Serie 'PKW-PKW'

Variable	T	df	p
korrekte Details gesamt	-3,708	21	0,001
korrekte Details im 'free recall'	-1,236	21	0,230
korrekte Details 'auf Nachfrage'	-3,165	21	0,005
Fehler gesamt	-0,799	21	0,433
Quellenverwechslungen insgesamt	1,003	21	0,327
Quellenverwechslungen innerhalb	-1,242	16,002	0,232
Quellenverwechslungen zwischen	3,105	10,000	0,011
schematypische Intrusionen	1,301	21	0,207
Konfabulationen	-1,603	21	0,124
allgemeine Erinnerungsfehler	-0,770	21	0,450

Dia- bzw. Unfall-Serie 'PKW-Fahrrad':

Bei gesonderter Analyse der Serie 'PKW-Fahrrad' ergab sich jedoch ein gänzlich anderes Bild als für die Dia-Serie 'PKW-PKW':

Hier erinnern bzw. berichten die Jugendlichen beider Versuchsgruppen insgesamt, im freien Bericht wie auch auf Nachfrage in etwa gleich viel korrekte Details. Es lässt sich somit kein signifikanter Unterschied beobachten derart, dass sich entweder an die Dia-Serie oder an die selbst erfundene Unfall-Serie besser oder schlechter erinnert wurde als an die andere.

Ebenso konnten für die Serie 'PKW-Fahrrad' keine Unterschiede zwischen den Gruppen bezüglich irgendeiner Art von Quellenverwechslung gefunden werden, auch nicht zwischen den Serien – im Gegensatz zur Abfrage von Dia-Serie 'PKW-PKW'. Bei Abfrage der Serie 'PKW-Fahrrad' unterlaufen den Versuchspersonen beider Bedingungen in etwa gleich viele oder wenige Quellenverwechslungen insgesamt, innerhalb oder zwischen den Serien, keine der Gruppen kann also hier besser oder schlechter zwischen und innerhalb der jeweiligen Quellen differenzieren.

Allerdings treten bei Befragung der selbst ausgedachten Unfall-Serie 'PKW-Fahrrad' im Vergleich zur Dia-Serie 'PKW-Fahrrad' deutlich mehr Fehler insgesamt ($T(21) = -3,398$ und $p = 0,003$), erkennbar mehr allgemeine Erinnerungsfehler ($T(21) = -2,193$ und $p = 0,040$) sowie auch wesentlich mehr Konfabulationen ($T(13,500) = -3,687$ und $p = 0,003$) auf. Das Erinnern an eine erfundene Unfall-Serie scheint hier fehleranfälliger zu sein als das Erinnern an eine Dia-Serie.

Ein mit $T(21) = 2,361$ und $p = 0,028$ statistisch ebenfalls signifikanter Unterschied findet sich hinsichtlich schematischer Intrusionen, wobei hier diesbezüglich die RM – Gruppe weniger Fehler begeht als die ESM – Gruppe.

Die nachstehenden *Tabellen 4-10 und 4-11* enthalten die jeweiligen Parameter und Testergebnisse:

Tabelle 4-10: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum ersten Befragungszeitpunkt T1 – für Serie 'PKW-Fahrrad'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 20,82; SD = 5,25	MW = 18,33; SD = 9,91
korrekte Details im 'free recall'	MW = 9,73; SD = 4,36	MW = 7,25; SD = 5,14
korrekte Details 'auf Nachfrage'	MW = 11,09; SD = 4,18	MW = 11,08; SD = 8,05
Fehler gesamt	MW = 12,45; SD = 3,21	MW = 17,50; SD = 3,85
Quellenverwechslungen insgesamt	MW = 0,82; SD = 0,87	MW = 1,00; SD = 0,95
Quellenverwechslungen innerhalb	MW = 0,18; SD = 0,41	MW = 0,50; SD = 0,67
Quellenverwechslungen zwischen	MW = 0,64; SD = 0,81	MW = 0,50; SD = 0,80
schematypische Intrusionen	MW = 1,09; SD = 0,83	MW = 0,42; SD = 0,52
Konfabulationen	MW = 0,55; SD = 0,69	MW = 2,92; SD = 2,11
allgemeine Erinnerungsfehler	MW = 10,18; SD = 2,96	MW = 13,17; SD = 3,51

Tabelle 4-11: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum ersten Befragungszeitpunkt T1 – für Serie 'PKW-Fahrrad'

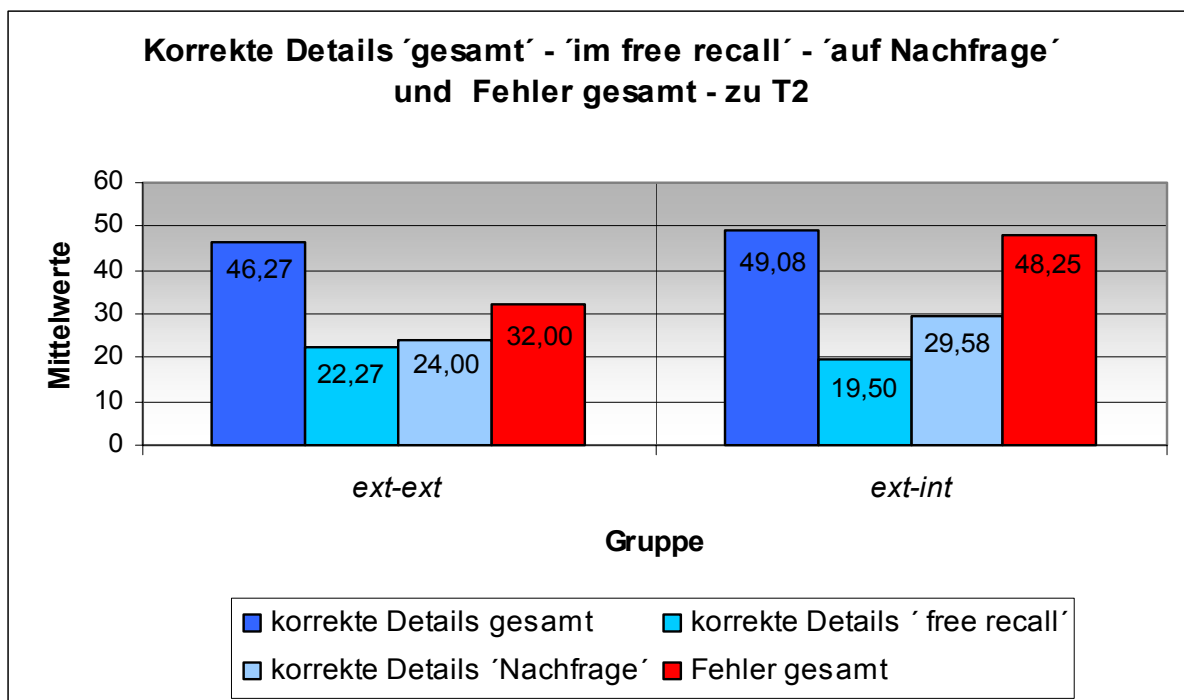
Variable	T	df	p
korrekte Details gesamt	0,760	17,011	0,458
korrekte Details im 'free recall'	1,241	21	0,228
korrekte Details 'auf Nachfrage'	0,003	21	0,998
Fehler gesamt	-3,398	21	0,003
Quellenverwechslungen insgesamt	-0,475	21	0,639
Quellenverwechslungen innerhalb	-1,385	18,242	0,183
Quellenverwechslungen zwischen	0,407	21	0,688
schematypische Intrusionen	2,361	21	0,028
Konfabulationen	-3,687	13,500	0,003
allgemeine Erinnerungsfehler	-2,193	21	0,040

4.1.2.2 Ergebnisse zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging

In diese Analyse werden pro Versuchsgruppe dieselben Versuchspersonen miteinbezogen, wie zum ersten Befragungszeitpunkt T1. Insgesamt handelt es sich auch hier wieder um $N = 23$ Versuchspersonen, wovon wiederum die Gruppe externe 'source monitoring' aus $n = 11$ und die 'reality monitoring' - Gruppe aus $n = 12$ Versuchspersonen bestand.

Diese zweite Befragung fand nach sieben bzw. acht Wochen statt.

Abbildung 4-4: Vergleich der beiden Gruppen hinsichtlich der korrekten Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und der insgesamt begangenen Fehler zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien



Zum zweiten Befragungszeitpunkt T2 zeigt sich bezüglich der berichteten korrekten Details ein ähnliches Bild wie zu T1. Auch hier berichtet die Gruppe 'external-internal' mehr korrekte Details insgesamt als auch mehr korrekte Details auf Nachfrage als die Gruppe 'external-external'. Lediglich im free recall berichtet die Gruppe 'external-internal' weniger korrekte Details als die Gruppe 'external-external'. All diese Unterschiede erweisen sich jedoch statistisch gesehen als auch zufällig möglich, sie sind also nicht signifikant ($T(21) = -0,466$; $p = 0,646$ und $T(21) = -1,327$; $p = 0,199$ bzw. $T(21) = 0,788$; $p = 0,440$).

Ein völlig anders Bild ergibt sich bei der Betrachtung der zu T2 insgesamt gemachten Fehler. Hier lässt sich bei der 'reality monitoring' - Gruppe eine sehr deutlich höhere Fehlerzahl

feststellen als bei der externalen 'source monitoring' - Gruppe. Der Unterschied in den Gruppenmittelwerten erweist sich als höchstsignifikant ($T(21) = -6,694$; $p < 0,001$).

Die einzelnen Parameter und Signifikanzergebnisse sind den nachstehenden Tabellen zu entnehmen.

Tabelle 4-12: Mittelwerte und Standardabweichungen der beiden Gruppen für diese Variablen zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien

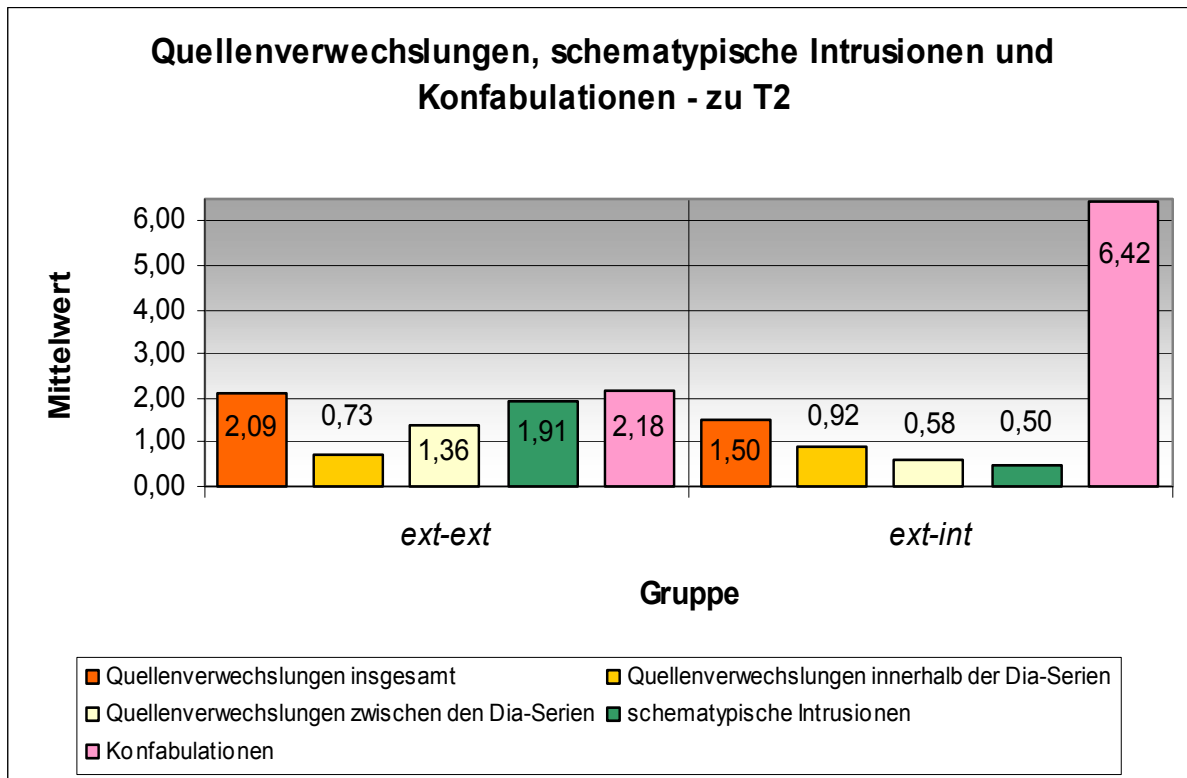
Gruppe	k. Det. 'gesamt'	k. Det. 'free recall'	k. Det. 'Nachfrage'	Fehler gesamt
ext-ext	MW = 46,27 SD = 12,11	MW = 22,27 SD = 7,80	MW = 24,00 SD = 10,19	MW = 32,00 SD = 6,07
ext-int	MW = 49,08 SD = 16,28	MW = 19,50 SD = 8,97	MW = 29,58 SD = 9,98	MW = 48,25 SD = 5,58

Tabelle 4-13: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien

Variable	T	df	Signifikanz
korrekte Details gesamt	-0,466	21	0,646
korrekte Details im 'free recall'	0,788	21	0,440
korrekte Details 'auf Nachfrage'	-1,327	21	0,199
Fehler gesamt	-6,694	21	< 0,001

Zum zweiten Befragungszeitpunkt T2 ergaben sich für den Fall einer schon vorangegangenen Befragung für die Variablen 'Quellenverwechslungen', 'schematypischen Intrusionen', 'Konfabulationen' und 'allgemeine Erinnerungsfehler' folgende Ergebnisse:

Abbildung 4-5: Vergleich der beiden Gruppen hinsichtlich der Quellenverwechslung insgesamt sowie beider Arten der Quellenwechslung, der schematypischen Intrusionen und der Konfabulationen zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien

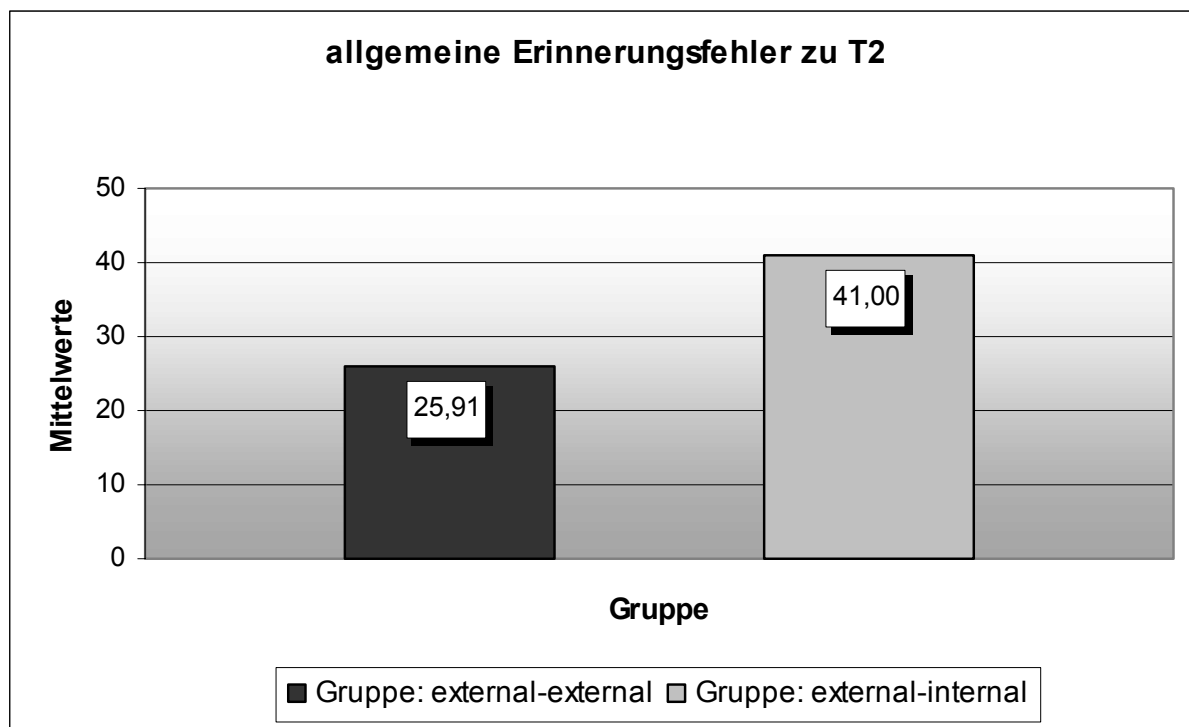


In der Bedingung 'external-internal' finden sich weniger Quellenverwechslungen insgesamt als in der Bedingung 'external-external'. Allerdings weist der t-Test dafür kein signifikantes Ergebnis aus ($T(21) = 0,937$; $p = 0,359$). Ebenso finden sich in der Gruppe 'external-internal' weniger Quellenverwechslungen zwischen den Dia-Serien, jedoch erweist sich auch dieser Unterschied als nicht-signifikant ($T(21) = 1,769$; $p = 0,091$). Bei den Quellenverwechslungen innerhalb der Dia-Serien steigen diese in der Gruppe 'external-internal' im Vergleich zu denen der Gruppe 'external-external' zwar leicht an, aber diese Veränderung bleibt statistisch ebenfalls nicht signifikant ($T(21) = -0,333$; $p = 0,743$).

Beim Vergleich der schematypischen Intrusionen erreicht die ESM-Gruppe hier einen um fast vierfach höheren Wert als die RM-Gruppe. Somit ergibt sich mit $T(21) = 3,467$; $p = 0,002$ ein hochsignifikantes Testergebnis.

Ein statistisch gleiches Ergebnis lässt sich beim Vergleich der Konfabulationen in den Gruppen. Jedoch ergibt sich hier für die 'reality monitoring' - Gruppe ein deutlich höherer Wert als für die externe 'source monitoring' - Gruppe. Demnach differenziert diese Variable zu T2 ebenfalls hochsignifikant zwischen den beiden Gruppen ($T(21) = -3,581$; $p = 0,002$).

Abbildung 4-6: Vergleich der beiden Gruppen hinsichtlich der allgemeinen Erinnerungsfehler zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien



Ein im Gruppenvergleich zu den Konfabulationen analoges Ergebnis zeigt sich bei der Analyse der allgemeinen Erinnerungsfehler. Die Versuchspersonen der Bedingung 'external-internal' machen hier im Mittel gut 14 Fehler mehr als die Versuchspersonen in der Bedingung 'external-external'. Mit einem höchstsignifikanten Ergebnis des t-Tests ($T(21) = -6,070$; $p < 0,001$) erweist sich die Variable 'allgemeine Erinnerungsfehler' als sehr guter Indikator, um zwischen den Gruppen zu trennen.

Im Anschluss sind wieder die konkreten Mittelwerte und Standardabweichungen sowie die Ergebnisse der t-Tests den *Tabellen 4-14* und *4-15* zu entnehmen.

Tabelle 4-14: Mittelwerte und Standardabweichungen der beiden Gruppen für diese Variablen zum T2, wenn schon eine Befragung vorausging – für beide Serien

Gruppe	QV total	QV intra	QV inter	Intrusionen	Konfabulation	AEF
ext-ext	MW = 2,09 SD = 1,70	MW = 0,73 SD = 1,27	MW = 1,36 SD = 1,29	MW = 1,91 SD = 1,04	MW = 2,18 SD = 2,27	MW = 25,91 SD = 5,21
ext-int	MW = 1,50 SD = 1,31	MW = 0,92 SD = 1,44	MW = 0,58 SD = 0,79	MW = 0,50 SD = 0,91	MW = 6,42 SD = 3,26	MW = 41,00 SD = 5,56

Tabelle 4-15: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen Quellenverwechslungen innerhalb und zwischen den Dia-Serien, schematypische Intrusionen, Konfabulationen und allgemeine Erinnerungsfehler zum zweiten Befragungszeitpunkt T2, wenn schon eine Befragung vorausging – für beide Serien

Variable	T	df	Signifikanz
Quellenverwechslungen insgesamt	0,937	21	0,359
Quellenverwechslungen innerhalb	-0,333	21	0,743
Quellenverwechslungen zwischen	1,769	21	0,091
schematypische Intrusionen	3,467	21	0,002
Konfabulationen	-3,581	21	0,002
allgemeine Erinnerungsfehler	-6,070	21	< 0,001

Betrachtet man nun die beiden Unfall-Serien getrennt voneinander ergibt sich folgendes Bild:

Dia-Serie 1 (PKW-PKW):

Hier zeigt sich hinsichtlich der korrekt berichteten Details das exakt gleiche Ergebnis wie zum ersten Befragungszeitpunkt. Insgesamt sowie auf Nachfrage berichten die Versuchspersonen der RM – Gruppe hochsignifikant mehr Details als diejenigen der ESM – Gruppe ($T(18,390) = -2,992$ und $p = 0,008$ bzw. $T(21) = -3,193$ und $p = 0,004$). Wiederum nicht-signifikant das Ergebnis des t-Tests für die korrekten Details im free recall, wenn auch hier sowie zu T1 die RM – Gruppe einen höheren Wert aufweist als die ESM – Gruppe.

Ein zum zweiten Befragungszeitpunkt T2 nach acht Wochen hingegen anderes Ergebnis (im Vergleich zu T1) zeigt sich nun jedoch hinsichtlich der Gesamtfehlerzahl sowie der Zahl der allgemeinen Erinnerungsfehler. Jetzt begehen die Versuchspersonen der RM – Gruppe bezüglich der Dia-Serie 'PKW-PKW' deutlich mehr Fehler als diejenigen der Gruppe ESM ($T(21) = -2,775$ und $p = 0,011$ bzw. $T(21) = -3,035$ und $p = 0,006$).

Wie auch zu T1 in dieser Dia-Serie ergaben sich für das Vorkommen schematypischer Intrusionen sowie von Konfabulationen keinerlei statistisch bedeutsamen, sprich signifikanten Unterschiede zwischen den Versuchsbedingungen.

Wiederum exakt analog zum ersten Befragungszeitpunkt T1 hinsichtlich der Dia-Serie 'PKW-PKW' wies der t-Test für die Quellenverwechslungen insgesamt und innerhalb der Serien kein signifikantes Ergebnis aus, jedoch ein solches für die Quellenverwechslungen zwischen den Serien ($T(10,000) = 2,667$ und $p = 0,024$). So kommen auch nach acht Wochen bei den

Jugendlichen der Bedingung 'external-internal' keinerlei Quellenverwechslungen zwischen den Serien vor. Diese können bei Befragung zur Dia-Serie 'PKW-PKW' also nach acht Wochen ebenso gut wie nach einer Woche zwischen den Quellen differenzieren und deutlich besser als die Versuchspersonen der Bedingung 'external-external'.

Nachstehenden *Tabellen 4-16 und 4-17* sind die genauen Werte zu entnehmen:

Tabelle 4-16: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum zweiten Befragungszeitpunkt T2 – für Serie 'PKW-PKW'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 25,18; SD = 6,59	MW = 36,25; SD = 10,81
korrekte Details im 'free recall'	MW = 11,64; SD = 3,72	MW = 13,75; SD = 7,06
korrekte Details 'auf Nachfrage'	MW = 13,55; SD = 6,55	MW = 22,50; SD = 6,87
Fehler gesamt	MW = 15,64; SD = 3,59	MW = 20,33; SD = 4,44
Quellenverwechslungen insgesamt	MW = 1,27; SD = 1,49	MW = 0,42; SD = 0,67
Quellenverwechslungen innerhalb	MW = 0,55; SD = 1,21	MW = 0,42; SD = 0,67
Quellenverwechslungen zwischen	MW = 0,73; SD = 0,91	MW = 0,00; SD = 0,00
schematypische Intrusionen	MW = 0,73; SD = 0,79	MW = 0,25; SD = 0,45
Konfabulationen	MW = 1,27; SD = 1,68	MW = 2,33; SD = 1,78
allgemeine Erinnerungsfehler	MW = 12,36; SD = 3,56	MW = 17,33; SD = 4,29

Tabelle 4-17: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum zweiten Befragungszeitpunkt T2 – für Serie 'PKW-PKW'

Variable	T	df	p
korrekte Details gesamt	-2,992	18,390	0,008
korrekte Details im 'free recall'	-0,885	21	0,386
korrekte Details 'auf Nachfrage'	-3,193	21	0,004
Fehler gesamt	-2,775	21	0,011
Quellenverwechslungen insgesamt	1,751	13,613	0,102
Quellenverwechslungen innerhalb	0,319	21	0,753
Quellenverwechslungen zwischen	2,667	10,000	0,024
schematypische Intrusionen	1,764	15,675	0,097
Konfabulationen	-1,469	21	0,157
allgemeine Erinnerungsfehler	-3,035	21	0,006

Dia- bzw. Unfall-Serie 'PKW-Fahrrad':

Die bezüglich korrekt berichteter Details zu T1 schon erkennbare, aber statistisch unbedeutende Tendenz findet zum zweiten Befragungszeitpunkt T2 nun zumindest für die Gesamtzahl als auch für die im freien Bericht erinnerten Details auch signifikanten Niederschlag. Die Jugendlichen der Gruppe RM erinnern bzw. berichten jetzt wesentlich weniger korrekte Details als diejenigen der Gruppe ESM. Es finden sich mit $T(21) = 2,298$ und $p = 0,032$ bzw. $T(21) = 2,426$ und $p = 0,024$ jeweils signifikante Testergebnisse. Der Unterschied 'auf Nachfrage' ist zwar auch hier gut erkennbar, jedoch nicht signifikant.

Zum ersten Befragungszeitpunkt T1 analoge Befunde in Form von mindestens hochsignifikanten Testergebnissen zeigten sich für wiederum die Gesamtfehlerzahl ($T(21) = -5,358$ und $p < 0,001$), die Rate der allgemeinen Erinnerungsfehler ($T(21) = -3,726$ und $p = 0,001$), schematypische Intrusionen ($T(21) = 2,967$ und $p = 0,007$) sowie Konfabulationen ($T(15,688) = -4,805$ und $p < 0,001$). Auch zum zweiten Befragungszeitpunkt T2 lag die entsprechende Fehlerrate in der RM –Gruppe deutlich höher als in der ESM – Gruppe mit Ausnahme des Befundes der schematypischen Intrusionen - wie auch zu T1 fanden sich in der Bedingung 'external-external' mehr Intrusionen.

Ebenso analog zu T1 konnten auch für die Serien 'PKW-Fahrrad' keinerlei signifikante Unterschiede zwischen den Versuchsgruppen, was Quellenverwechslungen insgesamt, innerhalb oder zwischen den Serien angeht, gefunden werden.

Näheres hierzu kann den *Tabellen 4-18 und 4-19* entnommen werden:

Tabelle 4-18: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum zweiten Befragungszeitpunkt T2 – für Serie 'PKW-Fahrrad'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 21,09; SD = 7,75	MW = 12,83; SD = 9,32
korrekte Details im 'free recall'	MW = 10,64; SD = 5,75	MW = 5,75; SD = 3,79
korrekte Details 'auf Nachfrage'	MW = 10,45; SD = 6,74	MW = 7,08; SD = 7,93
Fehler gesamt	MW = 16,36; SD = 4,50	MW = 29,17; SD = 6,65
Quellenverwechslungen insgesamt	MW = 0,82; SD = 0,98	MW = 1,17; SD = 0,94
Quellenverwechslungen innerhalb	MW = 0,18; SD = 0,60	MW = 0,58; SD = 0,90
Quellenverwechslungen zwischen	MW = 0,64; SD = 0,67	MW = 0,58; SD = 0,79
schematypische Intrusionen	MW = 1,18; SD = 0,87	MW = 0,25; SD = 0,62
Konfabulationen	MW = 0,91; SD = 0,94	MW = 4,08; SD = 2,07
allgemeine Erinnerungsfehler	MW = 13,55; SD = 4,63	MW = 23,67; SD = 7,83

Tabelle 4-19: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum zweiten Befragungszeitpunkt T2 – für Serie ‘PKW-Fahrrad’

Variable	T	df	p
korrekte Details gesamt	2,298	21	0,032
korrekte Details im ‘free recall’	2,426	21	0,024
korrekte Details ‘auf Nachfrage’	1,093	21	0,287
Fehler gesamt	-5,358	21	< 0,001
Quellenverwechslungen insgesamt	-0,871	21	0,394
Quellenverwechslungen innerhalb	-1,244	21	0,227
Quellenverwechslungen zwischen	0,172	21	0,865
schematypische Intrusionen	2,967	21	0,007
Konfabulationen	-4,805	15,688	< 0,001
allgemeine Erinnerungsfehler	-3,726	21	0,001

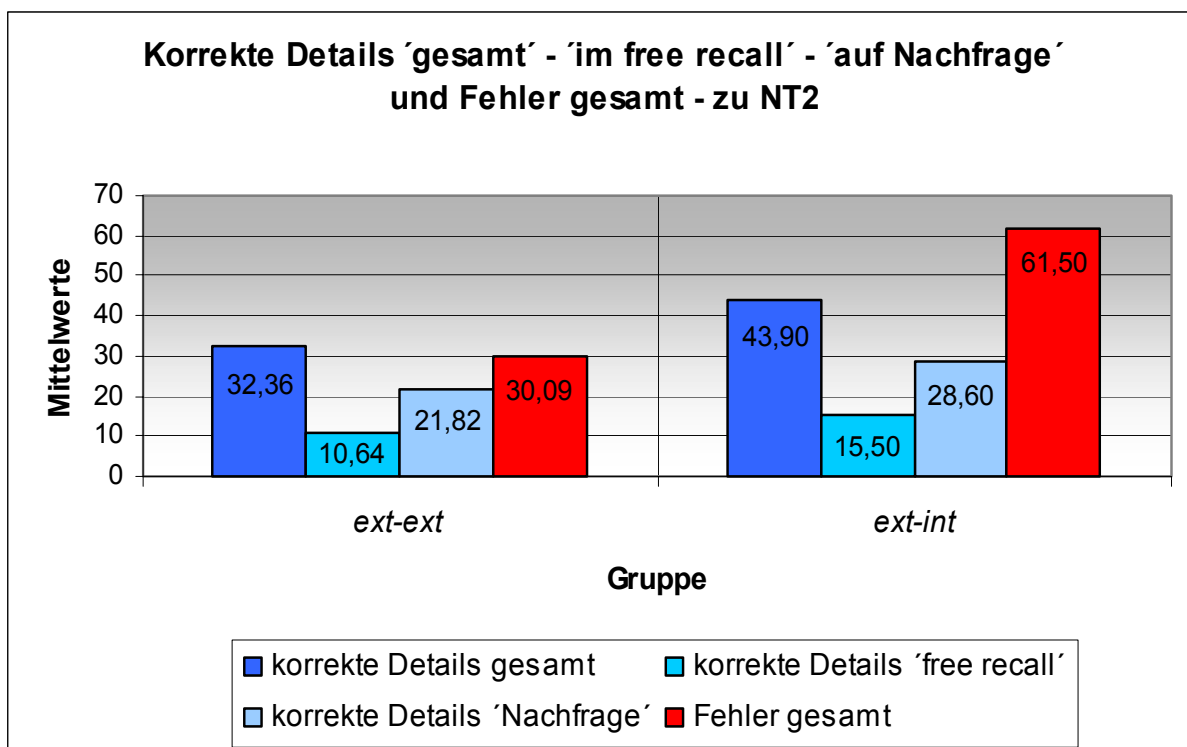
4.1.2.3 Ergebnisse zum zweiten (einzigen) Befragungszeitpunkt NT2, ohne eine vorausgehende Befragung

In diese Analyse wurden nun diejenigen Versuchspersonen miteinbezogen, für die nach sieben bzw. acht Wochen die erste und einzige Befragung stattfand, die also an keiner vorausgegangenen Befragung teilgenommen hatten. Diese Interviews fanden also zum selben Zeitpunkt statt, wie denjenigen für die Versuchspersonen, die schon an T1 teilgenommen hatten und auch zu T2 zum zweiten Mal befragt wurden.

Die Gesamtzahl der Versuchspersonen zu NT2 betrug für diese Teilgruppe $N = 21$. In der Bedingung 'external-external' lag die Teilstichprobengröße bei $n = 11$, in der Bedingung 'external-internal' bei $n = 10$.

Es ergaben sich bezüglich der abhängigen Variablen folgende Ergebnisse:

Abbildung 4-7: Vergleich der beiden Gruppen hinsichtlich der korrekten Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und der insgesamt begangenen Fehlern zum zweiten Befragungszeitpunkt NT2, ohne eine vorausgehende Befragung – für beide Serien



Sieht man sich die Ergebnisse dieser Teilstichprobe hinsichtlich der berichteten korrekten Details an, so ergibt sich ein schon bekanntes Bild. Auch zu diesem Befragungszeitpunkt berichten die Versuchspersonen der Gruppe 'external-internal' zwar einige korrekte Details

gesamt mehr als die der Gruppe 'external-external', ein signifikantes Ergebnis wird mit $T(19) = -1,904$; $p = 0,072$ allerdings auch hier wiederum nicht erzielt. Analog verhält es sich für sowohl die Details im freien Bericht als auch auf Nachfrage – auch hier weist der t-Test für unabhängige Stichproben jeweils keine signifikanten Ergebnisse aus ($T(19) = -1,876$; $p = 0,076$ bzw. $T(19) = -1,626$ und $p = 0,120$), obwohl die RM - Gruppe für beide Variablen doch sichtbar höhere Werte aufweist als die ESM - Gruppe.

Wie jedoch auch schon zu den beiden anderen Befragungszeitpunkten T1 und T2 lässt sich beim Vergleich der Gesamtfehlerzahlen ein im Ergebnis sehr klarer, jedoch bei Betrachtung der Mittelwerte noch deutlicherer Befund, festhalten. Die Versuchspersonen der 'reality monitoring' - Gruppe machen hier insgesamt beträchtlich mehr Fehler als die Versuchspersonen der anderen Gruppe, sie liegen im Mittel über 31 Fehlern darüber. Dieser Befund äußert sich dementsprechend in einem höchstsignifikanten Testergebnis ($T(19) = 0,678$; $p < 0,001$).

Siehe *Tabellen 4-20 und 4-21* für alle genauen Parameterangaben und Testwerte.

Tabelle 4-20: Mittelwerte und Standardabweichungen der beiden Gruppen für diese Variablen zum T2, ohne eine vorausgehende Befragung – für beide Serien

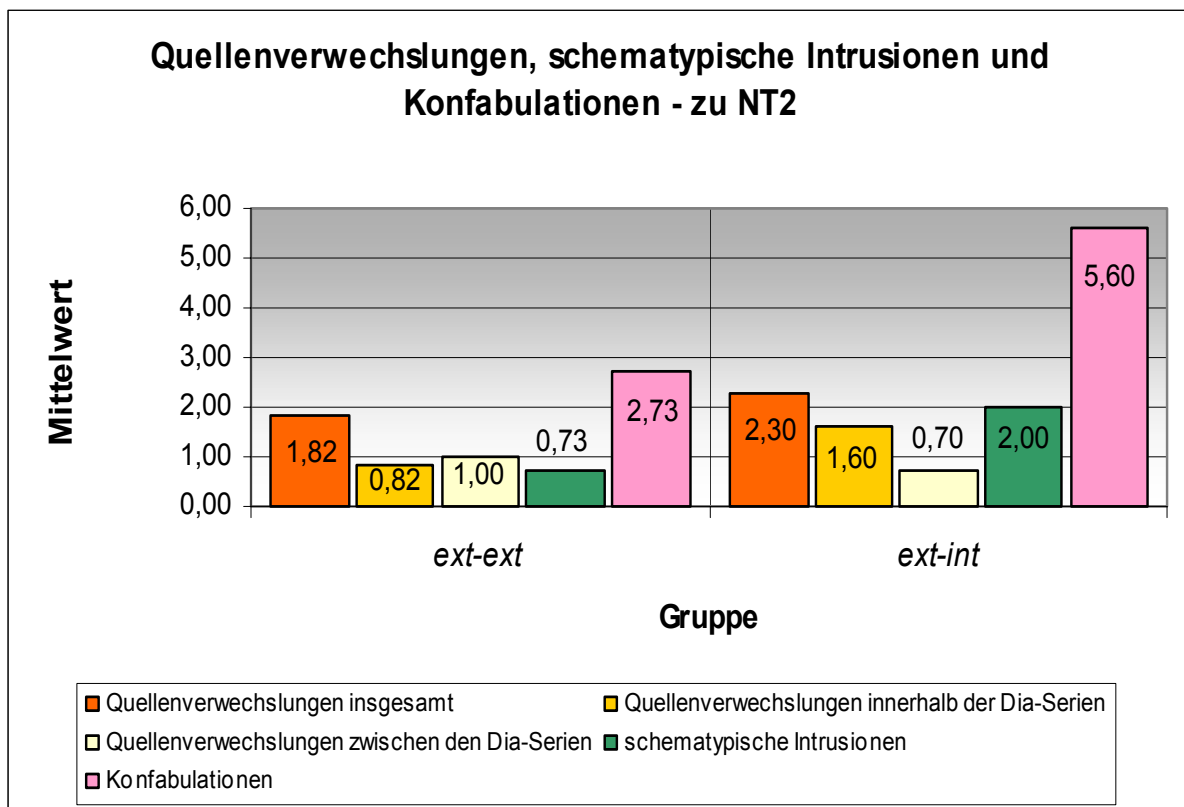
Gruppe	k. Det. 'gesamt'	k. Det. 'free recall'	k. Det. 'Nachfrage'	Fehler gesamt
ext-ext	MW = 32,36 SD = 13,55	MW = 10,64 SD = 5,75	MW = 21,82 SD = 9,01	MW = 30,09 SD = 6,98
ext-int	MW = 43,90 SD = 14,21	MW = 15,50 SD = 6,13	MW = 28,60 SD = 10,11	MW = 61,50 SD = 7,76

Tabelle 4-21: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zum zweiten Befragungszeitpunkt T2, ohne eine vorausgehende Befragung – für beide Serien

Variable	T	df	Signifikanz
korrekte Details gesamt	-1,904	19	0,072
korrekte Details im 'free recall'	-1,876	19	0,076
korrekte Details 'auf Nachfrage'	-1,626	19	0,120
Fehler gesamt	0,678	19	< 0,001

In Bezug auf die übrigen Fehlerarten wie Quellenverwechslungen, Intrusionen, Konfabulationen und allgemeine Erinnerungsfehler zeigen sich zu NT2 die nachfolgenden Befunde:

Abbildung 4-8: Vergleich der beiden Gruppen hinsichtlich der Quellenverwechslung insgesamt sowie beider Arten der Quellenwechslung, der schematypischen Intrusionen und der Konfabulationen zum zweiten Befragungszeitpunkt T2, ohne eine vorausgehende Befragung – für beide Serien



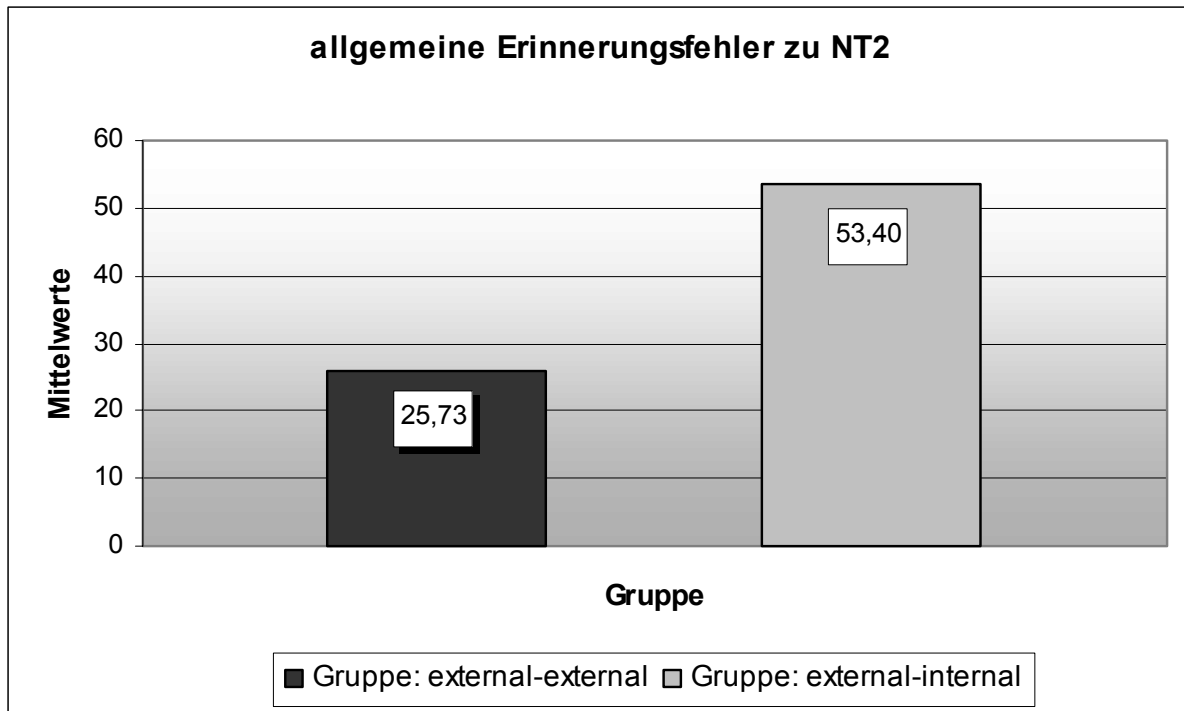
Für alle Arten der Quellenverwechslungen (gesamt, innerhalb und zwischen den Serien) ergaben sich hier keine signifikanten Mittelwertsunterschiede zwischen den beiden Versuchsbedingungen (siehe *Tabelle 4-23*). Bezüglich der Quellenverwechslungen gesamt und innerhalb der Dia-Serien lässt sich zumindest eine Tendenz feststellen, in der Form, dass die Mittelwerte der Gruppe 'external-internal' über denen der Gruppe 'external-external' liegen. Für die Quellenverwechslungen zwischen den Serien gilt allerdings das Umgekehrte.

Bei Betrachtung der schematypischen Intrusionen erreicht die RM-Gruppe einen erkennbar höheren Fehlerwert als die ESM-Gruppe. Dies schlägt sich in einem signifikanten Testergebnis nieder ($T(12,345) = -2,609$ und $p = 0,022$).

Auch in Bezug auf die Konfabulationen haben die Versuchspersonen der Bedingung 'external-internal' wesentlich größere Schwierigkeiten bei der Wiedergabe und der Erinnerungsleistung als diejenigen in der Bedingung 'external-external'. Sie begehen im Durchschnitt zu diesem Befragungszeitpunkt fast drei Konfabulationen mehr. Jedoch wird mit T

(11,271) = -1,978; $p = 0,073$ hier ein signifikanter Gruppenunterschied im Testergebnis verfehlt.

Abbildung 4-9: Vergleich der beiden Gruppen hinsichtlich der allgemeinen Erinnerungsfehler zum zweiten Befragungszeitpunkt T2, ohne eine vorausgehende Befragung – für beide Serien



Besonders deutlich wird der Unterschied zwischen den Versuchsgruppen bei den allgemeinen Erinnerungsfehlern. Die Gruppe 'external-internal' erinnert diesbezüglich sehr viel weniger als die Gruppe 'external-external'. Der Unterschied ist für diese Variable mit ca. 28 Fehlern der größte im Vergleich zu den beiden anderen Befragungszeitpunkten T1 und T2. Der t-Test weist demzufolge ein höchstsignifikantes Ergebnis aus ($T(19) = -7,457$; $p < 0,001$).

Für alle genauen Mittelwerte, Standardabweichungen und Testergebniswerte: siehe nachstehende *Tabellen 4-22 und 4-23*.

Tabelle 4-22: Mittelwerte und Standardabweichungen der beiden Gruppen für diese Variablen zum T2, ohne eine vorausgehende Befragung – für beide Serien

Gruppe	QV total	QV intra	QV inter	Intrusionen	Konfabulation	AEF
ext-ext	MW = 1,82 SD = 1,08	MW = 0,82 SD = 1,17	MW = 1,00 SD = 0,78	MW = 0,73 SD = 0,65	MW = 2,73 SD = 1,62	MW = 25,73 SD = 8,03
ext-int	MW = 2,30 SD = 2,06	MW = 1,60 SD = 1,27	MW = 0,70 SD = 1,06	MW = 2,00 SD = 1,41	MW = 5,60 SD = 4,33	MW = 53,40 SD = 8,98

Tabelle 4-23: Ergebnisse des t – Tests für unabhängige Stichproben für die Variablen Quellenverwechslungen innerhalb und zwischen den Dia-Serien, schematypische Intrusionen, Konfabulationen und allgemeine Erinnerungsfehler zum zweiten Befragungszeitpunkt NT2, ohne eine vorausgehende Befragung – für beide Serien

Variable	T	df	Signifikanz
Quellenverwechslungen insgesamt	-0,682	19	0,504
Quellenverwechslungen innerhalb	-1,473	19	0,157
Quellenverwechslungen zwischen	0,746	19	0,465
schematypische Intrusionen	-2,609	12,345	0,022
Konfabulationen	-1,978	11,271	0,073
allgemeine Erinnerungsfehler	-7,457	19	< 0,001

Betrachtet man nun die beiden Unfall-Serien getrennt voneinander ergibt sich folgendes Bild:

Dia-Serie 1 (PKW-PKW):

Die schon von T1 und T2 bekannte Tendenz der jeweils immer höheren Werte der RM – Gruppe gegenüber der ESM – Gruppe in den Dia-Serien 'PKW-PKW' hinsichtlich aller drei Variablen 'korrekte Details' (gesamt, free recall und auf Nachfrage) kann auch bei erstmaliger Befragung nach acht Wochen (NT2) beobachtet werden. Jedoch ergaben sich nur zu diesem Befragungszeitpunkt *für alle drei* signifikante t-Test-Ergebnisse. Nur zu NT2 sind also auch die Unterschiede zwischen den Versuchsbedingungen hinsichtlich der im 'free recall' berichteten korrekten Details auch statistisch bedeutsam. Es finden sich insgesamt ein hochsignifikantes Ergebnis ($T(19) = -2,916$ und $p = 0,009$), im freien Bericht und auf Nach-

frage jeweils signifikante Unterschiede ($T(19) = -2,470$ und $p = 0,023$ bzw. $T(19) = -2,720$ und $p = 0,014$).

Ein für die Dia-Serie 'PKW-PKW' zu T2 analoger Befund ergab sich bezüglich der insgesamt begangenen Fehler sowie der allgemeinen Erinnerungsfehler. Auch hier machen die Jugendlichen der Bedingung 'external-internal' nach acht Wochen höchst- bzw. hochsignifikant mehr Fehler ($T(19) = -3,988$ und $p = 0,001$ bzw. $T(19) = -3,445$ und $p = 0,003$) als diejenigen der Bedingung 'external-external' - also unabhängig davon, ob zu diesem Zeitpunkt eine erste oder bereits eine zweite Befragung stattfand.

Ein im statistischen Ergebnis zu T2 ebenso analoges Bild konnte hinsichtlich schematischer Intrusionen und auch Konfabulationen festgestellt werden. Bei erstmaliger Befragung nach acht Wochen fanden sich auch hier keine signifikanten Ergebnisse, allerdings mit dem tendenziellen Unterschied, dass zu NT2 die Versuchspersonen der RM – Gruppe mehr schematische Intrusionen berichten als die der ESM – Gruppe.

Ein sogar zu T1 und T2 analoges, besonderes Ergebnis findet sich wiederum hinsichtlich der zwischen den Serien begangenen Quellenverwechslungen. Bei Abfrage der Dia-Serie 'PKW-PKW' unterlaufen den Versuchspersonen der RM – Gruppe wie zu T1 und zu T2 auch zu NT2 keinerlei Quellenverwechslungen zwischen den Serien. Der diesbezügliche Unterschied zur ESM – Gruppe ist mit $T(10,000) = 2,631$ und $p = 0,025$ auch hier wieder signifikant. Die Versuchspersonen der RM – Gruppe sind gegenüber denjenigen der ESM – Gruppe demnach sehr gut bzw. besser in der Lage zwischen den Serien und Quellen zu differenzieren, egal nach welcher Zeit und wie oft sie dazu befragt wurden. Für die Unterschiede der Quellenverwechslungen innerhalb der Serien sowie insgesamt konnten ebenso wieder keine statistisch bedeutsamen Unterschiede beobachtet werden.

Aus den *Tabellen 4-24 und 4-25* können die dazu gehörigen Mittelwerte, Standardabweichungen, T-, df- und p-Werte abgelesen werden:

Tabelle 4-24: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum zweiten Befragungszeitpunkt NT2 – für Serie 'PKW-PKW'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 20,27; SD = 9,23	MW = 34,10; SD = 12,41
korrekte Details im 'free recall'	MW = 6,36; SD = 3,72	MW = 11,80; SD = 6,18
korrekte Details 'auf Nachfrage'	MW = 13,91; SD = 6,02	MW = 22,30; SD = 8,06
Fehler gesamt	MW = 16,45; SD = 4,01	MW = 24,40; SD = 5,10
Quellenverwechslungen insgesamt	MW = 1,00; SD = 0,78	MW = 0,90; SD = 0,99
Quellenverwechslungen innerhalb	MW = 0,45; SD = 0,82	MW = 0,90; SD = 0,99
Quellenverwechslungen zwischen	MW = 0,55; SD = 0,69	MW = 0,00; SD = 0,00
schematypische Intrusionen	MW = 0,55; SD = 0,69	MW = 1,10; SD = 0,74
Konfabulationen	MW = 1,55; SD = 0,82	MW = 2,70; SD = 1,89
allgemeine Erinnerungsfehler	MW = 13,36; SD = 3,93	MW = 19,70; SD = 4,50

Tabelle 4-25: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum zweiten Befragungszeitpunkt NT2 – für Serie 'PKW-PKW'

Variable	T	df	p
korrekte Details gesamt	-2,916	19	0,009
korrekte Details im 'free recall'	-2,470	19	0,023
korrekte Details 'auf Nachfrage'	-2,720	19	0,014
Fehler gesamt	-3,988	19	0,001
Quellenverwechslungen insgesamt	0,258	19	0,799
Quellenverwechslungen innerhalb	-1,124	19	0,275
Quellenverwechslungen zwischen	2,631	10,000	0,025
schematypische Intrusionen	-1,783	19	0,091
Konfabulationen	-1,786	12,033	0,099
allgemeine Erinnerungsfehler	-3,445	19	0,003

Dia- bzw. Unfall-Serie 'PKW-Fahrrad':

Im statistischen Ergebnis gleichen die folgenden Befunde zu NT2 (nach acht Wochen) exakt den Befunden zu T1, also ebenfalls einer ersten Befragung – nach einer Woche. Es ist lediglich in der Richtungstendenz der schematypischen Intrusionen ein Unterschied zu verzeichnen.

Zunächst kann hinsichtlich der korrekt berichteten Details (insgesamt, im freien Bericht und auf Nachfrage) festgestellt werden, dass hier zwischen den Gruppen kein signifikanter Unterschied besteht, obwohl die RM – Gruppe wie auch zu T1 und zu T2 in Bezug auf ihre selbst erfundene Unfall-Serie tendenziell weniger Details berichtet als die ESM – Gruppe bezüglich der Dia-Serie 'PKW-Fahrrad'.

Der schon von den T1 und T2 bekannte Befund, dass die Versuchspersonen der Bedingung 'external-internal' bei Schilderung der selbst erfundenen Unfall-Serie 'PKW-Fahrrad' deutlich mehr Fehler insgesamt sowie im Bereich der allgemeinen Erinnerungsfehler begehen als die Versuchspersonen der Bedingung 'external-external' lässt sich auch zu NT2 nicht nur feststellen, die angesprochenen Unterschiede sind jetzt jeweils sogar mit Abstand am größten. Dies äußert sich in jeweils höchstsignifikanten Ergebnissen des t-Tests für unabhängige Stichproben ($T(13,541) = -8,420$ und $p < 0,001$ bzw. $T(11,377) = -7,053$ und $p < 0,001$).

Wie auch zu den beiden anderen Befragungszeitpunkten konnten bei Abfrage dieser Serie 'PKW-Fahrrad' hinsichtlich irgendeiner Art von Quellenverwechslung innerhalb oder zwischen den Serien keinerlei signifikante Unterschiede zwischen den beiden Versuchsbedingungen gefunden werden.

Konfabuliert wird innerhalb der Serie 'PKW-Fahrrad' auch bei einer ersten Befragung nach acht Wochen (NT2) wie auch nach einer Woche (T1) sowie ebenso bei einer zweiten Befragung nach acht Wochen (T2) hingegen wiederum deutlich mehr in der RM – Gruppe als in der ESM – Gruppe, was sich erneut in einem signifikanten Ergebnis des t-Tests widerspiegelt ($T(13,998) = -2,278$ und $p = 0,038$).

Der bereits angedeutete Richtungsunterschied hinsichtlich des Vorkommens schematischer Intrusionen fällt offenbar nun ins Gewicht. Zwar ergibt sich hier ebenfalls mit $T(19) = -2,207$ und $p = 0,040$ ein signifikanter Unterschied zwischen den Versuchsgruppen für die Serie 'PKW-Fahrrad', jedoch äußern im Gegensatz zu T1 und T2 nun die Jugendlichen der 'external-internen' Bedingung *weniger* Intrusionen als diejenigen der 'external-externen' Bedingung.

Die eben dargestellten Ergebnisse können durch die nachfolgenden *Tabellen 4-26 und 4-27* ergänzt und detaillierter betrachtet werden:

Tabelle 4-26: Mittelwerte und Standardabweichungen der beiden Gruppen für alle Variablen zum zweiten Befragungszeitpunkt NT2 – für Serie 'PKW-Fahrrad'

Variable	Gruppe	
	ESM	RM
korrekte Details gesamt	MW = 12,18; SD = 5,83	MW = 9,80; SD = 5,25
korrekte Details im 'free recall'	MW = 4,27; SD = 2,90	MW = 3,50; SD = 3,21
korrekte Details 'auf Nachfrage'	MW = 7,91; SD = 4,11	MW = 6,30; SD = 3,80
Fehler gesamt	MW = 13,64; SD = 4,59	MW = 39,10; SD = 8,50
Quellenverwechslungen insgesamt	MW = 0,82; SD = 1,33	MW = 1,40; SD = 1,65
Quellenverwechslungen innerhalb	MW = 0,36; SD = 0,92	MW = 0,70; SD = 0,82
Quellenverwechslungen zwischen	MW = 0,45; SD = 0,69	MW = 0,70; SD = 1,06
schematypische Intrusionen	MW = 0,18; SD = 0,41	MW = 0,90; SD = 0,99
Konfabulationen	MW = 1,18; SD = 1,33	MW = 3,10; SD = 2,33
allgemeine Erinnerungsfehler	MW = 11,45; SD = 3,59	MW = 33,70; SD = 9,37

Tabelle 4-27: Ergebnisse des t – Tests für unabhängige Stichproben für alle Variablen zum zweiten Befragungszeitpunkt NT2 – für Serie 'PKW-Fahrrad'

Variable	T	df	p
korrekte Details gesamt	0,981	19	0,339
korrekte Details im 'free recall'	0,580	19	0,569
korrekte Details 'auf Nachfrage'	0,928	19	0,365
Fehler gesamt	-8,420	13,541	< 0,001
Quellenverwechslungen insgesamt	-0,895	19	0,382
Quellenverwechslungen innerhalb	-0,877	19	0,392
Quellenverwechslungen zwischen	-0,636	19	0,532
schematypische Intrusionen	-2,207	19	0,040
Konfabulationen	-2,278	13,998	0,038
allgemeine Erinnerungsfehler	-7,053	11,377	< 0,001

Abschließend folgt eine tabellarische Zusammenfassung aller t-Test-Ergebnisse zu den Gruppenunterschieden zu den verschiedenen Befragungszeitpunkten – zunächst wiederum für beide Serien zusammen sowie für beide Serien getrennt voneinander.

Tabelle 4-28: zusammenfassende Darstellung der Ergebnisse aller t – Tests für alle Variablen getrennt nach Befragungszeitpunkten für die externale ‘source-monitoring’ - Gruppe vs. die ‘reality-monitoring’ - Gruppe – für beide Serien

Variable	Befragungszeitpunkt		
	T1	T2	NT2
korrekte Details gesamt			
korrekte Details im ‘free recall’			
korrekte Details ‘auf Nachfrage’	*		
Fehler gesamt	*	***	***
Quellenverwechslungen insgesamt			
Quellenverwechslungen innerhalb			
Quellenverwechslungen zwischen	*		
schematypische Intrusionen	*	**	*
Konfabulationen	**	**	
allgemeine Erinnerungsfehler		***	***

Tabelle 4-29: zusammenfassende Darstellung der Ergebnisse aller t – Tests für alle Variablen getrennt nach Befragungszeitpunkten für die externale ‘source-monitoring’ - Gruppe vs. die ‘reality-monitoring’ - Gruppe – für Serie ‘PKW-PKW’

Variable	Befragungszeitpunkt		
	T1	T2	NT2
korrekte Details gesamt	***	**	**
korrekte Details im ‘free recall’			*
korrekte Details ‘auf Nachfrage’	**	**	*
Fehler gesamt		*	***
Quellenverwechslungen insgesamt			
Quellenverwechslungen innerhalb			
Quellenverwechslungen zwischen	*	*	*
schematypische Intrusionen			
Konfabulationen			
allgemeine Erinnerungsfehler		**	**

Tabelle 4-30: zusammenfassende Darstellung der Ergebnisse aller t – Tests für alle Variablen getrennt nach Befragungszeitpunkten für die externale 'source-monitoring' - Gruppe vs. die 'reality-monitoring' - Gruppe – für Serie 'PKW-Fahrrad'

Variable	Befragungszeitpunkt		
	T1	T2	NT2
korrekte Details gesamt		*	
korrekte Details im 'free recall'		*	
korrekte Details 'auf Nachfrage'			
Fehler gesamt	**	***	***
Quellenverwechslungen insgesamt			
Quellenverwechslungen innerhalb			
Quellenverwechslungen zwischen			
schematypische Intrusionen	*	**	*
Konfabulationen	**	***	*
allgemeine Erinnerungsfehler	*	***	***

wobei: ' ' = kein signifikantes Ergebnis ($p > 0,05$)

'*' = signifikantes Ergebnis ($p \leq 0,05$)

'**' = hochsignifikantes Ergebnis ($p \leq 0,01$)

'***' = höchstsignifikantes Ergebnis ($p \leq 0,001$)

Tabelle 4-31: Zusammenfassende Darstellung der Unterschiede zwischen ESM- und RM – Gruppe zu allen Befragungszeitpunkten (T1, T2 und NT2) für beide Unfall-Serien zusammen sowie getrennt voneinander

Variable	Befragungszeitpunkt								
	T1			T2			NT2		
	PKW-PKW. PKW-Fahrrad	PKW- PKW	PKW- Fahrrad	PKW-PKW. PKW-Fahrrad	PKW- PKW	PKW- Fahrrad	PKW-PKW. PKW-Fahrrad	PKW- PKW	PKW- Fahrrad
korrekte Details gesamt		***			**	*		**	
korr. Details im 'free recall'						*		*	
korr. Details 'auf Nachfrage'	*	**			**			*	
Fehler gesamt	*		**	***	*	***	***	***	***
QV insgesamt									
QV innerhalb									
QV zwischen	*	*			*			*	
schematypische Intrusionen	*		*	**		**	*		*
Konfabulationen	**		**	**		***			*
allgem. Erinnerungsfehler			*	***	**	***	***	**	***

wobei: ' ' = kein signifikantes Ergebnis ($p > 0,05$) ' * ' = signifikantes Ergebnis ($p \leq 0,05$)

' ** ' = hochsignifikantes Ergebnis ($p \leq 0,01$) ' *** ' = höchstsignifikantes Ergebnis ($p \leq 0,001$)

4.1.3 Ergebnisse der Veränderung der Gedächtnisleistung über die Zeit

Im Folgenden werden die Ergebnisse der Veränderung der Gedächtnisleistungen beider Gruppen getrennt voneinander über die Zeit hinweg dargestellt.

In diese Analyse werden also diejenigen Versuchspersonen der externalen 'source monitoring' sowie der 'reality monitoring' - Gruppe miteinbezogen, die sowohl zum ersten Befragungszeitpunkt T1 als auch zum zweiten Befragungszeitpunkt T2 interviewt wurden.

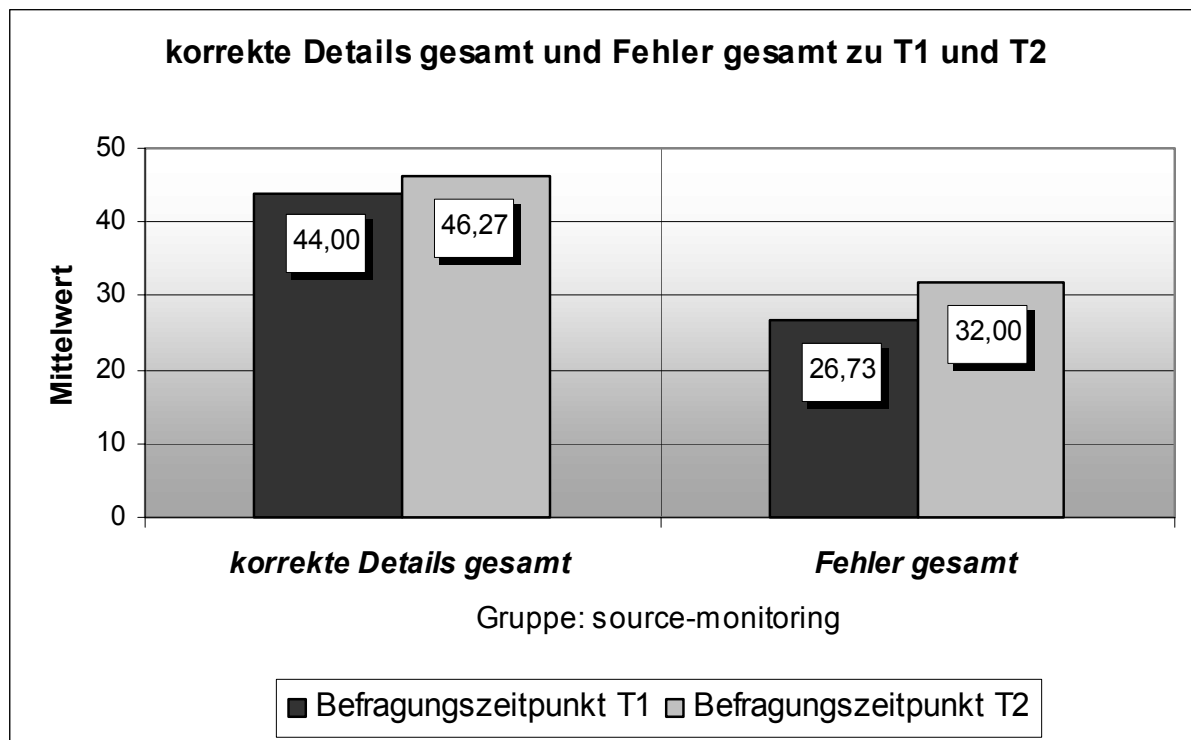
Insgesamt betrug die Stichprobengröße $N = 22$, $n = 11$ für jede der beiden Versuchsgruppen. Zur Analyse der Gruppen-Mittelwertsunterschiede von T1 nach T2 wurde jeweils ein t-Test für abhängige Stichproben durchgeführt.

Die Ergebnisse werden für die beiden Gruppen separat betrachtet und erläutert.

4.1.3.1 Ergebnisse für die externe 'source monitoring' – Gruppe

Den Versuchspersonen dieser Gruppe ($n = 11$) wurden beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad' dargeboten. Es ergaben sich folgende Ergebnisse:

Abbildung 4-10: Darstellung der insgesamt berichteten korrekten Details und der Fehler gesamt zu den beiden Befragungszeitpunkten für die externe 'source monitoring' - Gruppe. Die Angaben beziehen sich auf beide Dia-Serien.

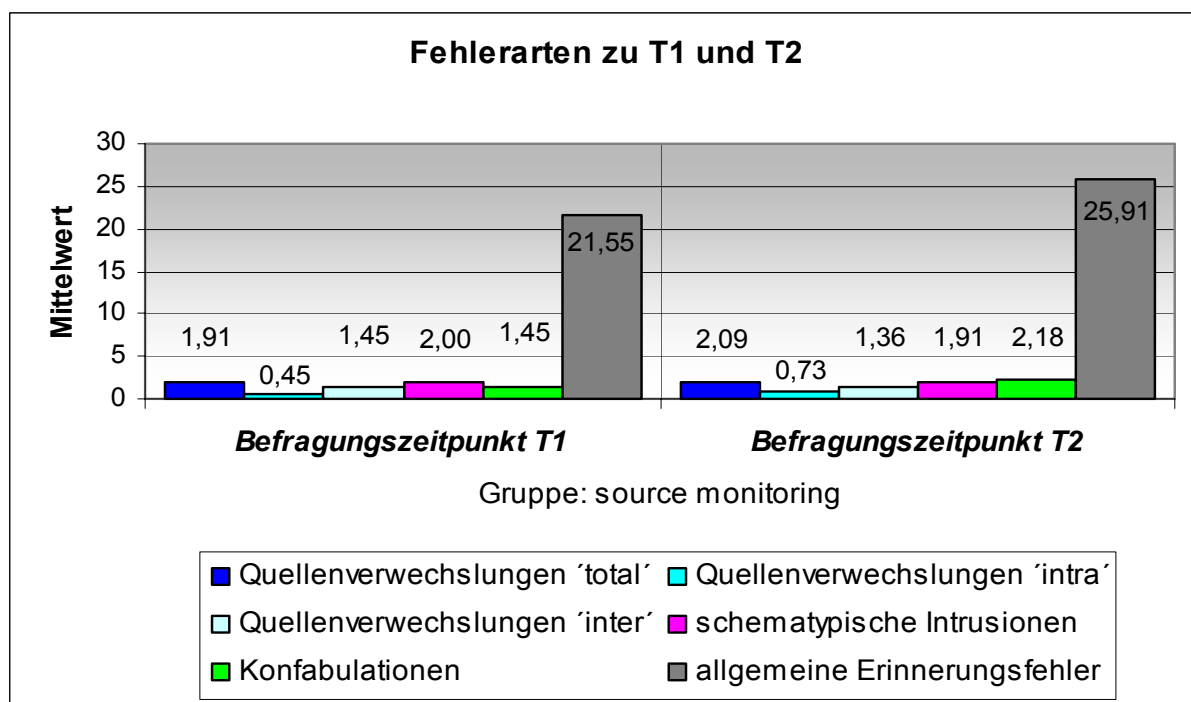


Entgegen der Annahme steigt der Anteil der insgesamt berichteten korrekten Details vom ersten Befragungszeitpunkt zum zweiten hin leicht an. Allerdings erweist sich dieser geringe Anstieg statistisch als nicht bedeutsam. Der Unterschied zwischen den Befragungszeitpunkten ist nicht signifikant ($T(10) = -0,680$; $p = 0,512$). Als ebenso gering erweisen sich die Unterschiede bezüglich der korrekten Details im free recall und auch auf Nachfrage. Beide Ergebnisse sind nicht signifikant.

Erwartungsgemäß und hypothesenkonform gestaltet sich das Ergebnis hinsichtlich der insgesamt begangenen Fehler in dieser Versuchsbedingung. Hier steigt die Fehlerrate vom ersten zum zweiten Befragungszeitpunkt höchstsignifikant an ($T(10) = -4,097$; $p = 0,001$).

Der Vergleich der beiden Befragungszeitpunkte bezüglich der einzelnen Fehlerarten zeichnet folgendes Bild:

Abbildung 4-11: Darstellung der einzelnen Fehlerarten zu beiden Befragungszeitpunkten für die externe 'source monitoring' - Gruppe. Die Angaben beziehen sich auf beide Dia-Serien.



Vorweg lässt sich zunächst feststellen, dass mit Ausnahme der allgemeinen Erinnerungsfehler sich bei keiner Fehlerart eine signifikante Veränderung zwischen den Befragungszeitpunkt finden lässt.

Es zeigt sich bei den Quellenverwechslungen insgesamt und innerhalb der Dia-Serien zwar jeweils ein leichter Anstieg von T1 nach T2, jedoch sind diese Unterschiede nicht signifikant ($T(10) = -0,276$; $p = 0,394$ bzw. $T(10) = -0,896$; $p = 0,196$). Hingegen werden zu T2 weniger

Quellenverwechslungen zwischen den Serien begangen als zu T1. Der Unterschied ist aber so gering, dass sich auch hier kein signifikantes Testergebnis findet ($T(10) = 0,191$; $p = 0,426$).

Ebenfalls ein leichter Abfall vom ersten zum zweiten Befragungszeitpunkt ist bei den schematypischen Intrusionen zu beobachten. Wie allerdings bei fast allen Variablen dieser Kategorie besteht hier ein so geringer Unterschied, dass sich dies in keinem signifikanten Ergebnis darstellt ($T(10) = 0,184$; $p = 0,429$).

Wiederum ein erwartungsgemäßer Anstieg über die Zeit lässt sich bei den Konfabulationsfehlern feststellen. Diese Veränderung bleibt jedoch in so geringem Ausmaß, dass der t-Test für abhängige Stichproben mit $T(10) = -1,018$ und $p = 0,167$ hier ebenfalls kein signifikantes Ergebnis bringt.

Hypothesenkonform und auch statistisch bedeutsam dagegen stellt sich das Ergebnis für die allgemeinen Erinnerungsfehler dar. Es kann hier ein signifikanter Anstieg in der Ausprägung dieser Variable vom ersten zum zweiten Befragungszeitpunkt konstatiert werden ($T(10) = -2,891$; $p = 0,008$). Die Versuchspersonen der ESM-Gruppe erinnern hier zu T2 deutlich weniger als zu T1.

Zusammenfassend ergeben sich somit für die externe 'source monitoring' - Gruppe lediglich bei der Gesamtzahl der Fehler sowie den allgemeinen Erinnerungsfehlern statistisch signifikante Veränderungen über die Zeit – bezogen auf beide Dia-Serien.

Die dazugehörigen statistischen Kennwerte sind den *Tabellen 4-32 und 4-33* zu entnehmen.

Tabelle 4-32: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu beiden Befragungszeitpunkten für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	44,00	8,19	46,27	12,11
korrekte Details im 'free recall'	18,55	7,56	22,27	7,80
korrekte Details 'auf Nachfrage'	25,45	5,99	24,00	10,19
Fehler gesamt	26,73	6,13	32,00	6,07
Quellenverwechslungen insgesamt	1,91	1,64	2,09	1,70
Quellenverwechslungen innerhalb	0,45	0,69	0,73	1,27
Quellenverwechslungen zwischen	1,45	1,29	1,36	1,29
schematypische Intrusionen	2,00	1,41	1,91	1,04
Konfabulationen	1,45	0,93	2,18	2,27
allgemeine Erinnerungsfehler	21,55	6,74	25,91	5,21

Tabelle 4-33: Ergebnisse des t – Tests für abhängige Stichproben für alle Variablen über die Zeit für die externale ‘source-monitoring’ – Gruppe. Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	-0,680	10	0,512
korrekte Details im ‘free recall’	-1,499	10	0,083
korrekte Details ‘auf Nachfrage’	0,458	10	0,329
Fehler gesamt	-4,097	10	0,001
Quellenverwechslungen insgesamt	-0,276	10	0,394
Quellenverwechslungen innerhalb	-0,896	10	0,196
Quellenverwechslungen zwischen	0,191	10	0,426
schematypische Intrusionen	0,184	10	0,429
Konfabulationen	-1,018	10	0,167
allgemeine Erinnerungsfehler	-2,891	10	0,008

Wie schon erwähnt bezogen sich oben stehende Ausführungen auf die Betrachtung beider Dia-Serien zusammen. Differenziert man nun die Analyse der berichteten korrekten Details und der begangenen Fehler nach beiden Dia-Serien getrennt, werden also für alle abhängigen Variablen die t-Tests für beide Serien getrennt berechnet, so ergeben sich für die externe ‘source monitoring - Gruppe pro Dia-Serie folgende Befunde:

Dia-Serie 1 (PKW-PKW):

Vorweg lässt sich zunächst feststellen, dass für keine der Variablen ein signifikantes Testergebnis gefunden wurde. Die Veränderungen über die Zeit in dieser Gruppe sind für die Serie ‘PKW-PKW’ somit durchgängig sehr gering.

Jedoch ist tendenziell für folgende Variablen ein *leichter Anstieg in der Ausprägung des jeweiligen Mittelwerts* vom ersten zum zweiten Befragungszeitpunkt zu verzeichnen:

- korrekte Details gesamt und im free recall
- Fehler gesamt
- Quellenverwechslungen gesamt und innerhalb der Dia-Serie
- Konfabulationen
- allgemeine Erinnerungsfehler

Ein *leichter Abfall im Mittelwert* fand sich hingegen bei folgenden Variablen:

- korrekte Details auf Nachfrage
- Quellenverwechslungen zwischen den Serien
- schematypische Intrusionen

Die genauen Mittelwerte, Standardabweichungen und auch die Ergebnisse der t-Tests sind nachstehend tabellarisch aufgeführt und zusammengefasst (*Tabellen 4-34 und 4-35*).

Tabelle 4-34: Mittelwerte und Standardabweichungen aller abhängigen Variablen zum ersten und zum zweiten Befragungszeitpunkt für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 1 (PKW-PKW).

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	23,18	6,27	25,18	6,59
korrekte Details im 'free recall'	8,82	3,87	11,64	3,72
korrekte Details 'auf Nachfrage'	14,36	5,99	13,55	6,55
Fehler gesamt	14,27	5,14	15,64	3,59
Quellenverwechslungen insgesamt	1,09	1,04	1,27	1,49
Quellenverwechslungen innerhalb	0,27	0,47	0,55	1,21
Quellenverwechslungen zwischen	0,82	0,87	0,73	0,91
schematypische Intrusionen	0,91	0,83	0,73	0,79
Konfabulationen	0,91	0,94	1,27	1,68
allgemeine Erinnerungsfehler	11,36	5,28	12,36	3,56

Dia-Serie 2 (PKW-Fahrrad):

Ein relativ ähnliches Bild ergibt sich für die zweite Dia-Serie 'PKW-Fahrrad'. Hier lässt sich lediglich für die Variablen 'Fehler gesamt' und 'allgemeine Erinnerungsfehler' eine bedeutsame Veränderung der Gedächtnisleistung über die Zeit in Form eines hochsignifikanten bzw. signifikanten Testergebnisses beobachten.

Erwartungsgemäß und hypothesenkonform nehmen die beiden Fehlerraten (Fehler gesamt und allgemeine Erinnerungsfehler) von T1 nach T2 zu. Belege dafür sind die Ergebnisse der t-Tests ($T(10) = -2,952$; $p = 0,007$ und $T(10) = -2,121$; $p = 0,030$).

Für alle übrigen abhängigen Variablen lassen sich in der Bedingung 'external-external' und für diese Dia-Serie keine signifikanten Veränderungen über die beiden Befragungszeitpunkte feststellen.

Tendenziell weisen folgende Variablen einen *leichten Anstieg* von T1 nach T2 auf:

- korrekte Details gesamt und im free recall
- schematypische Intrusionen
- Konfabulationen

Eine Veränderung im Sinne eines *leichten Abfalls* von T1 nach T2 ergibt sich nur für die Variable 'korrekte Details auf Nachfrage'.

Für alle Arten der Quellenverwechslungen (gesamt, innerhalb und zwischen der Serie) finden sich überhaupt keine *Veränderungen über die Zeit*. Die Mittelwerte bleiben identisch.

Die genauen Mittelwerte, Standardabweichungen und auch die Ergebnisse der t-Tests sind wiederum nachstehend tabellarisch aufgeführt und zusammengefasst (Tabellen 4-35 und 4-36).

Tabelle 4-35: Mittelwerte und Standardabweichungen aller abhängigen Variablen zum ersten und zum zweiten Befragungszeitpunkt T2 für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 2 (PKW-Fahrrad).

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	20,82	5,25	21,09	7,75
korrekte Details im 'free recall'	9,73	4,36	10,64	5,75
korrekte Details 'auf Nachfrage'	11,09	4,18	10,45	6,74
Fehler gesamt	12,45	3,21	16,36	4,50
Quellenverwechslungen insgesamt	0,82	0,87	0,82	0,98
Quellenverwechslungen innerhalb	0,18	0,41	0,18	0,60
Quellenverwechslungen zwischen	0,64	0,81	0,64	0,67
schematypische Intrusionen	1,09	0,83	1,18	0,87
Konfabulationen	0,55	0,69	0,91	0,94
allgemeine Erinnerungsfehler	10,18	2,96	13,55	4,63

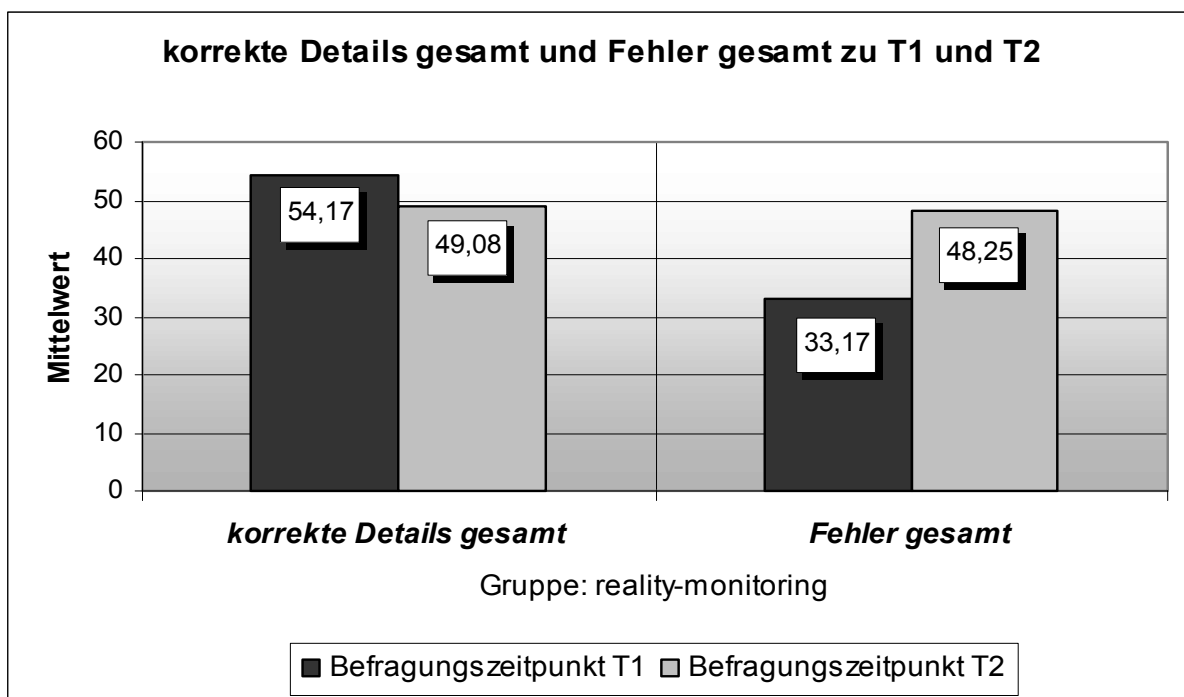
Tabelle 4-36: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Dia-Serien für die externe 'source-monitoring'-Gruppe über die Zeit (T1-T2)

Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	P
korrekte Details gesamt	-0,676	10	0,257	-0,170	10	0,435
korrekte Details im 'free recall'	-1,609	10	0,070	-0,609	10	0,278
korrekte Details 'auf Nachfrage'	0,258	10	0,401	0,319	10	0,378
Fehler gesamt	-1,000	10	0,171	-2,952	10	0,007
Quellenverwechslungen gesamt	-0,516	10	0,309	0,000	10	0,500
Quellenverwechslungen innerhalb	-0,896	10	0,196	0,000	10	0,500
Quellenverwechslungen zwischen	0,319	10	0,378	0,000	10	0,500
schematypische Intrusionen	0,614	10	0,277	-0,319	10	0,378
Konfabulationen	-0,841	10	0,210	-1,000	10	0,171
allgemeine Erinnerungsfehler	0,717	10	0,245	-2,121	10	0,030

4.1.3.2 Ergebnisse für die 'reality monitoring' – Gruppe

Den Versuchspersonen dieser Gruppe ($n = 11$) wurde zunächst die Dia-Serie 'PKW-PKW' dargeboten. Im Anschluss daran sollten sie sich anhand bestimmter Vorgaben und Anhaltspunkte einen zweiten Verkehrsunfall ähnlich bzw. der Dia-Serienform 'PKW-Fahrrad' selbst ausdenken. Es ergaben sich nach der Befragung dazu folgende Ergebnisse:

Abbildung 4-12: Darstellung der insgesamt berichteten korrekten Details und der Fehler gesamt zu den beiden Befragungszeitpunkten für die 'reality monitoring' - Gruppe. Die Angaben beziehen sich auf beide Unfall-Serien.

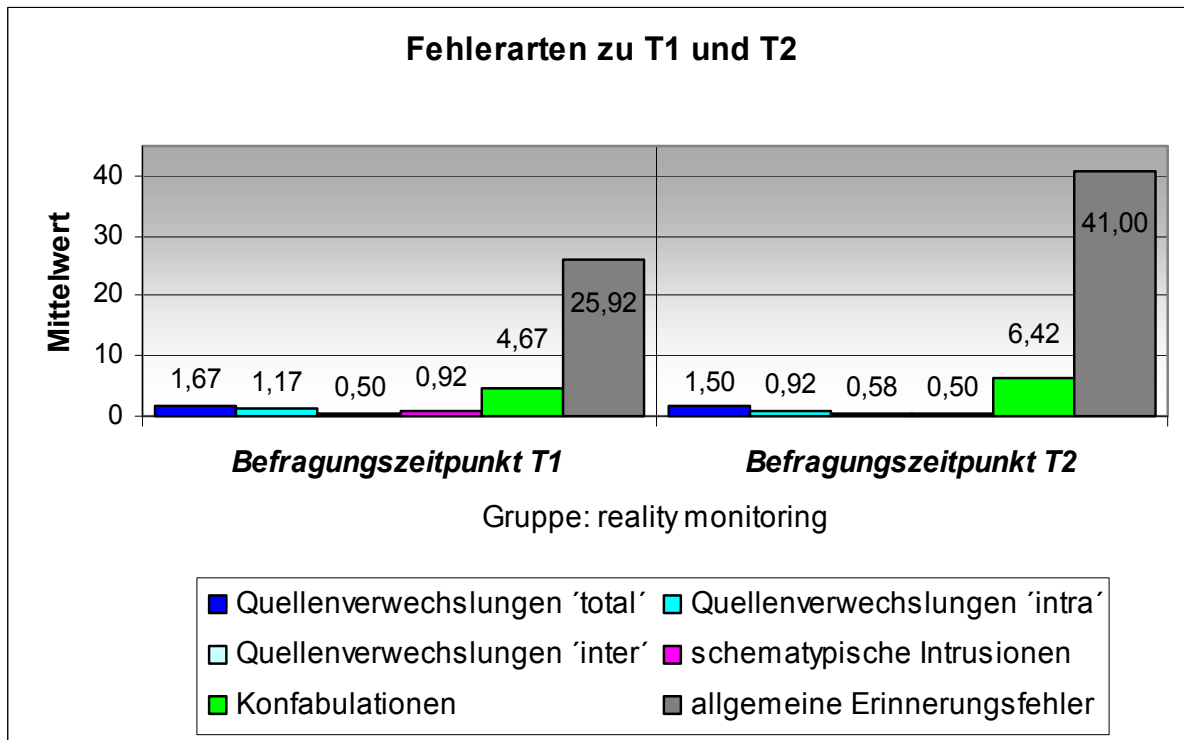


Die statistische Überprüfung der Veränderung der insgesamt korrekt berichteten Details über die Zeit zeigt einen signifikanten Abfall des Mittelwerts vom ersten zum zweiten Befragungszeitpunkt ($T(11) = 1,941$; $p = 0,039$). Ebenfalls signifikant ist die Veränderung über die Zeit bezüglich der korrekten Details auf Nachfrage ($T(11) = 2,243$; $p = 0,023$). Bei den korrekten Details im freien Bericht hingegen ist ein leichter Anstieg von T1 nach T2 zu beobachten. Dieser Unterschied ist jedoch so gering, dass dies statistisch nicht ins Gewicht fällt – das Ergebnis des t-Tests fällt nicht-signifikant aus.

Beim Vergleich der Gesamtfehlerzahl ist analog dazu ein starker Anstieg beim zweiten Befragungszeitpunkt festzustellen. In dieser Versuchsgruppe werden wie in der ESM-Gruppe zu T2 wesentlich mehr Fehler gemacht als zu T1. Hier ist der Unterschied sogar noch größer (mit über 15 Fehlern fast dreimal so groß), er lässt sich ebenfalls in einem höchstsignifikanten Testergebnis festhalten ($T(11) = -11,404$; $p < 0,001$).

Beim Vergleich der beiden Befragungszeitpunkte bezüglich der einzelnen Fehlerarten zeichnet sich für diese Versuchsgruppe folgendes Bild ab:

Abbildung 4-13: Darstellung der einzelnen Fehlerarten zu beiden Befragungszeitpunkten für die 'reality monitoring' - Gruppe. Die Angaben beziehen sich auf beide Unfall-Serien.



Die Veränderungen über die Zeit hinsichtlich aller Arten der Quellenverwechslungen (gesamt, innerhalb und zwischen den Dia-Serien) sowie auch der schematypischen Intrusionen gestalten sich – wie auch in Abbildung 4-13 zu sehen – allesamt so gering, dass der t-Test für keine dieser Variablen ein signifikantes Ergebnis ausweist. Es kommen sowohl ein leichtes Ansteigen als auch ein leichtes Abfallen zwischen den Zeitpunkten der einzelnen Variablen vor.

Bei Betrachtung der Konfabulationen ist dagegen eine gut erkennbare und auch signifikante Veränderung vom ersten zum zweiten Befragungszeitpunkt vorhanden. Bei im Mittel 1,75 Konfabulationen mehr zu T2 als zu T1, wird auch das Ergebnis des t-Tests signifikant ($T(11) = -2,273$; $p = 0,022$).

Ein deutlich sichtbarer und ebenso statistisch bedeutsamer Unterschied zwischen den beiden Befragungszeitpunkten lässt sich ein weiteres Mal für die allgemeinen Erinnerungsfehler nachweisen. In dieser, der RM-Gruppe, erinnern die Versuchspersonen zu T2 sehr viel weniger Details und Elemente der Dia-Serie bzw. machen diesbezüglich deutlich mehr

Fehler als zu T1. Ein höchstsignifikantes Ergebnis mit $T(11) = -9,584$ und $p < 0,001$ bringt dies zum Ausdruck.

Wie in der gesamten bisherigen Darstellung der Ergebnisse üblich finden sich tabellarisch die konkreten Werte und Testergebnisse in den nachstehenden *Tabellen 4-37 und 4-38*.

Tabelle 4-37: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu beiden Befragungszeitpunkten für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf beide Unfall-Serien.

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	54,17	17,49	49,08	16,28
korrekte Details im 'free recall'	18,42	9,69	19,50	8,97
korrekte Details 'auf Nachfrage'	35,75	13,90	29,58	9,98
Fehler gesamt	33,17	5,73	48,25	5,58
Quellenverwechslungen insgesamt	1,67	1,44	1,50	1,31
Quellenverwechslungen innerhalb	1,17	1,03	0,92	1,44
Quellenverwechslungen zwischen	0,50	0,80	0,58	0,79
schematypische Intrusionen	0,92	0,79	0,50	0,91
Konfabulationen	4,67	2,81	6,42	3,26
allgemeine Erinnerungsfehler	25,92	5,25	41,00	6,56

Tabelle 4-38: Ergebnisse des t – Tests für abhängige Stichproben für alle Variablen über die Zeit für die 'reality-monitoring' – Gruppe. Die Angaben beziehen sich auf beide Unfall-Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	1,941	11	0,039
korrekte Details im 'free recall'	-0,572	11	0,290
korrekte Details 'auf Nachfrage'	2,243	11	0,023
Fehler gesamt	-11,404	11	< 0,001
Quellenverwechslungen insgesamt	0,312	11	0,381
Quellenverwechslungen innerhalb	0,506	11	0,312
Quellenverwechslungen zwischen	-0,432	11	0,337
schematypische Intrusionen	1,603	11	0,069
Konfabulationen	-2,273	11	0,022
allgemeine Erinnerungsfehler	-9,584	11	< 0,001

Wie für die ESM-Gruppe bereits dargestellt, folgen auch hier für die RM-Gruppe die Ergebnisse und Ausführungen der Veränderungen der Gedächtnisleistungen über die Zeit für die beiden Unfall-Serien getrennt voneinander. Es zeigten sich folgende Ergebnisse:

Dia-Serie 1 (PKW-PKW):

Bei Betrachtung der korrekt berichteten Details weisen die t-Tests für keine der drei Variablen ('gesamt', 'free recall' und auf 'Nachfrage') ein signifikantes Ergebnis aus $T(11) = -0,164$; $p = 0,437$, $T(11) = -1,734$; $p = 0,056$ und $T(11) = 1,015$; $p = 0,166$). Die diesbezüglichen Veränderungen über die Zeit bewegen sich also in recht kleinen Ausmaßen. Kritisch anzumerken ist allerdings die hypothesenkonträre Tatsache, dass zumindest die Mittelwerte der korrekten Details gesamt sowie der korrekten Details im free recall zu T2 höher sind als zu T1.

Was die Gesamtzahl der Fehler betrifft, so zeigt sich hier wieder ein erwartungsgemäßes Ergebnis. Die Fehlerrate ist zum zweiten Befragungszeitpunkt deutlich höher als zum ersten. Es lässt sich ein hochsignifikanter Anstieg über die Zeit festhalten ($T(11) = -3,561$; $p = 0,002$).

Bezüglich der Quellenverwechslungen ergeben sich keinerlei Veränderungen über die Zeit, geschweige denn signifikante. Die Mittelwerte für die Quellenverwechslungen insgesamt und innerhalb der Dia-Serie bleiben völlig identisch, ebenso zwischen den Dia-Serien. Hier ist kein Signifikanztest möglich, da die Mittelwerte und Standardabweichungen jeweils 0 betragen, also keine derartigen Quellenverwechslungen vorkommen.

Weiterhin findet sich kein nennenswerter Unterschied beim Vergleich der Konfabulationen über die Zeitpunkte hinweg. Das Ergebnis des t-Tests ist nicht-signifikant.

Anders stellt sich dies bei den schematypischen Intrusionen sowie den allgemeinen Erinnerungsfehlern dar. Beide Testergebnisse sind signifikant bzw. hochsignifikant. Während bei den Intrusionen ein signifikanter Abfall von T1 nach T2 zu beobachten ist ($T(11) = 1,915$; $p = 0,041$), steigen die allgemeinen Erinnerungsfehler von T1 nach T2 hochsignifikant an ($T(11) = -3,697$; $p = 0,002$).

Die *Tabellen 4-39* und *4-41* konkretisieren und ergänzen die eben ausgeführten Befunde.

Tabelle 4-39: Mittelwerte und Standardabweichungen aller abhängigen Variablen zum ersten und zum zweiten Befragungszeitpunkt für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf die erste Dia-Serie 'PKW-PKW'.

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	35,83	9,58	36,25	10,81
korrekte Details im 'free recall'	11,17	5,10	13,75	7,06
korrekte Details 'auf Nachfrage'	24,67	8,47	22,50	6,87
Fehler gesamt	15,67	3,06	20,33	4,44
Quellenverwechslungen insgesamt	0,67	0,99	0,42	0,67
Quellenverwechslungen innerhalb	0,67	0,99	0,42	0,67
Quellenverwechslungen zwischen	0,00	0,00	0,00	0,00
schematypische Intrusionen	0,50	0,67	0,25	0,45
Konfabulationen	1,75	1,49	2,33	1,78
allgemeine Erinnerungsfehler	12,75	3,19	17,33	4,23

Serie 2 (PKW-Radfahrer) - selbst ausgedachter Unfall – analog Dia-Serie 'PKW-Fahrrad':

Hier ergibt sich für die berichteten korrekten Details ein klares, zu erwartendes Bild. Alle drei Teilvariablen (korrekte Details 'gesamt', 'free recall' und 'auf Nachfrage') weisen zum zweiten Befragungszeitpunkt geringere Mittelwerte auf als zum ersten. Für die Gesamtzahl der korrekt berichteten Details sowie für die korrekten Details auf Nachfrage stellt sich dies in einem statistisch hochsignifikanten ($T(11) = 3,102$; $p = 0,005$) bzw. signifikanten ($T(11) = 2,111$; $p = 0,029$) Ergebnis dar. Für die im free recall korrekt berichteten Details kann kein signifikantes Ergebnis festgestellt werden ($T(11) = 1,129$; $p = 0,142$). Die Versuchspersonen sind also über die Zeit hinweg kaum bzw. schlecht in der Lage, die selbst ausgedachten Details zu erinnern und zu reproduzieren.

Der Vergleich der Gesamtfehlerzahl unterstützt indirekt bzw. analog dazu das gerade beschriebene Ergebnis für die korrekten Details. Zum zweiten Befragungszeitpunkt steigt der Mittelwert der insgesamt gemachten Fehler stark an und führt damit zu einem höchstsignifikanten Ergebnis im t-Test ($T(11) = -6,535$; $p < 0,001$).

Hinsichtlich aller Arten von Quellenverwechslungen finden sich auch für diese Versuchsgruppe für die selbst ausgedachte Unfall-Serie keinerlei signifikante Veränderungen vom ersten zum zweiten Befragungszeitpunkt.

Ebenfalls ist für die schematypischen Intrusionen keine signifikante Veränderung über die Zeit feststellbar. Die Intrusionen nehmen zu T2 allerdings leicht ab.

Anderes gilt für die zeitliche Veränderung von berichteten Konfabulationen. Hier steigt der Mittelwert von T1 nach T2 signifikant an ($T(11) = -2,310$; $p = 0,021$). Die RM-Gruppe neigt also bei der selbst ausgedachten Unfall-Serie dazu, zum zweiten Befragungszeitpunkt mehr zu konfabulieren als zum ersten.

Komplettiert werden die bisherigen Ausführungen diesbezüglich durch das Ergebnis des t-Tests für die Variable der allgemeinen Erinnerungsfehler. Offensichtlich sind die Versuchspersonen sehr schwer in der Lage, die selbst erfundenen Aspekte und Beschreibungen von Personen und Details zu T1 und T2 wie auch über die Zeit hinweg zu erinnern. Bei über zehn Fehlern mehr zum zweiten Befragungszeitpunkt gegenüber dem ersten lässt sich ein höchstsignifikanter Unterschied nachweisen ($T(11) = -5,631$; $p < 0,001$).

Mittelwerte, Standardabweichungen und die Ergebnisse des t-Tests für abhängige Stichproben sind im Folgenden tabellarisch dargestellt (*Tabelle 4-40 und 4-41*).

Tabelle 4-40: Mittelwerte und Standardabweichungen aller abhängigen Variablen zum ersten und zum zweiten Befragungszeitpunkt für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf Serie 2 (PKW-Fahrrad) – selbst ausgedachter Unfall - analog Dia-Serie 'PKW-Fahrrad'.

Variable	T1		T2	
	MW	SD	MW	SD
korrekte Details gesamt	18,33	9,91	12,83	9,32
korrekte Details im 'free recall'	7,25	5,14	5,75	3,79
korrekte Details 'auf Nachfrage'	11,08	8,05	7,08	7,93
Fehler gesamt	17,50	3,85	29,17	6,65
Quellenverwechslungen insgesamt	1,00	0,95	1,17	0,94
Quellenverwechslungen innerhalb	0,50	0,67	0,58	0,90
Quellenverwechslungen zwischen	0,50	0,80	0,58	0,79
schematypische Intrusionen	0,42	0,52	0,25	0,62
Konfabulationen	2,92	2,11	4,08	2,07
allgemeine Erinnerungsfehler	13,17	3,51	23,67	7,83

Tabelle 4-41: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Unfall-Serien für die 'reality-monitoring' – Gruppe über die Zeit (T1-T2)

Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	P
korrekte Details gesamt	-0,164	11	0,437	3,102	11	0,005
korrekte Details im 'free recall'	-1,734	11	0,056	1,129	11	0,142
korrekte Details 'auf Nachfrage'	1,015	11	0,166	2,111	11	0,029
Fehler gesamt	-3,561	11	0,002	-6,535	11	< 0,001
Quellenverwechslungen gesamt	0,761	11	0,232	-0,561	11	0,293
Quellenverwechslungen innerhalb	0,761	11	0,232	-0,290	11	0,389
Quellenverwechslungen zwischen	<i>n. berechenbar: MWe = 0</i>			-0,432	11	0,337
schematypische Intrusionen	1,915	11	0,041	0,804	11	0,219
Konfabulationen	-1,134	11	0,141	-2,310	11	0,021
allgemeine Erinnerungsfehler	-3,697	11	0,002	-5,631	11	< 0,001

4.1.4 Ergebnisse zu den Befragungszeitpunkten

4.1.4.1 Befragungszeitpunkt T1 (nach einer Woche) vs. Befragungszeitpunkt NT2 (nach sieben bzw. acht Wochen zum ersten Mal)

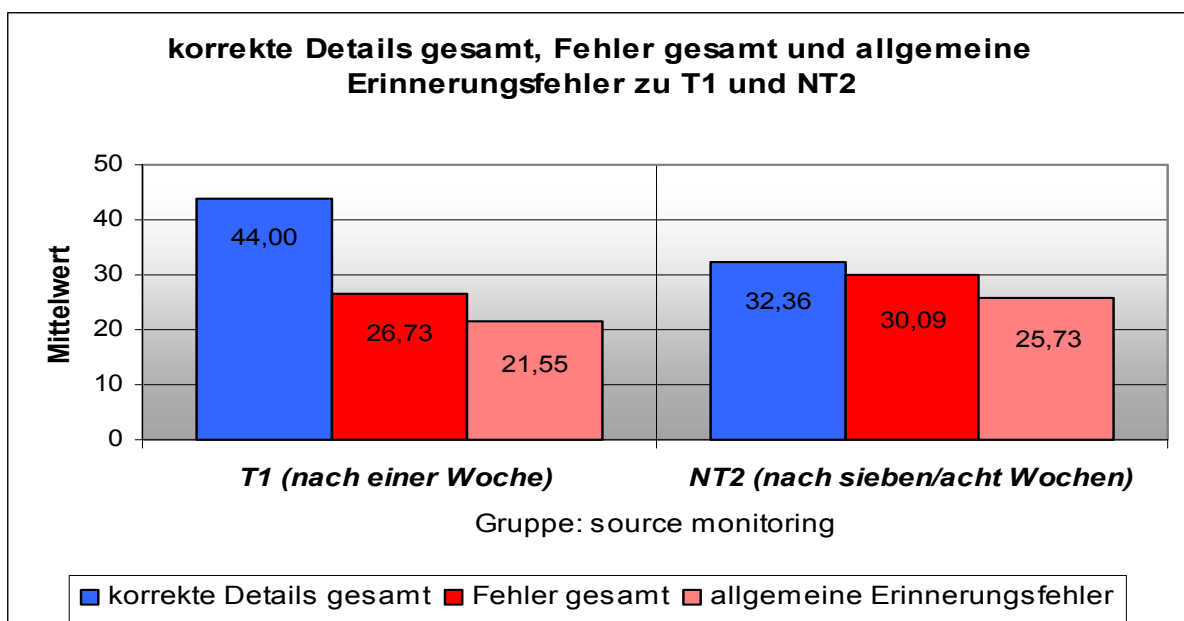
Ebenfalls praxisrelevant etwa im forensischen und juristischen Kontext und somit von besonderem Interesse ist die Frage, ob und inwieweit sich bei einer Opfer- oder Zeugenbefragung Unterschiede in der Erinnerungsleistung zeigen, wenn eine erste Befragung schon eine Woche oder erst sieben oder acht Wochen nach dem Erleben oder der Wahrnehmung eines Geschehens stattfindet.

Aus diesem Grund werden nun für beide Versuchsgruppen getrennt voneinander die Mittelwerte aller abhängigen Variablen zu den beiden genannten Befragungszeitpunkten mittels des t-Tests für unabhängige Stichproben miteinander verglichen (T1 vs. NT2).

4.1.4.1.1 Ergebnisse für die externe 'source monitoring' – Gruppe

Nach einer Woche (T1) wurden $n = 11$ Versuchspersonen interviewt, nach sieben bzw. acht Wochen (NT2) ebenfalls $n = 11$. Es ergaben sich folgende Ergebnisse:

Abbildung 4-14: Darstellung der insgesamt berichteten korrekten Details, der insgesamt berichteten Fehler und der allgemeinen Erinnerungsfehler für die externe 'source monitoring' – Gruppe bei einer ersten Befragung nach einer (T1) bzw. nach sieben oder acht Wochen (NT2). Die Angaben beziehen sich auf beide Dia-Serien.



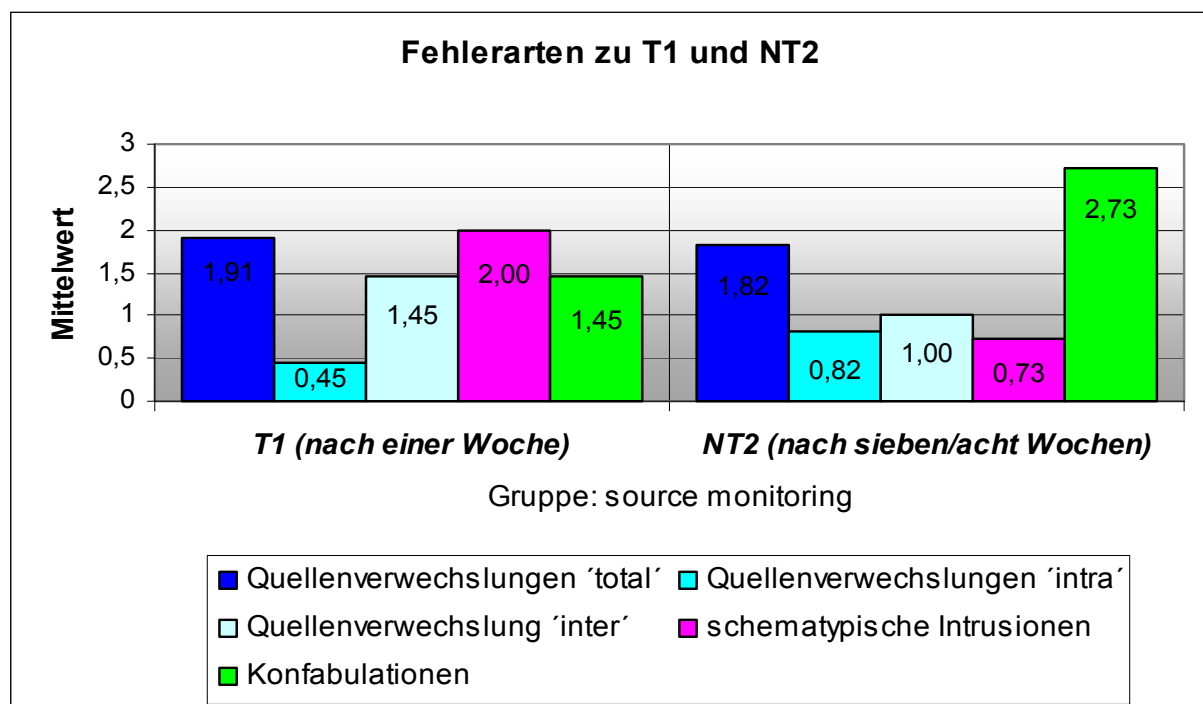
Bei Betrachtung der insgesamt korrekt berichteten Details ergibt sich mit $T(20) = -2,438$; $p = 0,012$) ein deutlicher und signifikanter - sogar fast hochsignifikanter – Unterschied zwischen den Befragungszeitpunkten. Werden die Versuchspersonen das erste Mal nach acht Wochen befragt, erinnern sie wesentlich weniger korrekte Details. Dieser Abfall zeigt sich noch deutlicher bei genauerer Analyse der korrekt berichteten Details im freien Bericht. Hier lässt sich sogar ein hochsignifikanter Unterschied feststellen ($T(20) = -2,763$; $p = 0,006$). Auf Nachfrage werden zu NT2 zwar ebenfalls weniger Details berichtet, jedoch verfehlt der Test hier das Signifikanzniveau von $p = 0,05$.

Tendenziell machen die Versuchspersonen bei einer ersten Befragung nach ca. acht Wochen auch insgesamt mehr Fehler gegenüber einer ersten Befragung nach bereits einer Woche. Dieser Anstieg der Gesamtfehlerzahl ist allerdings nicht signifikant ($T(20) = 1,201$; $p = 0,122$).

Analog verhält es sich hinsichtlich der allgemeinen Erinnerungsfehler. Auch hier finden sich nach ca. acht Wochen (NT2) mehr Fehler als bei erstmaliger Befragung nach einer Woche (T1). Der t-Test weist aber auch hier kein signifikantes Ergebnis aus ($T(20) = 1,323$; $p = 0,101$).

Hinsichtlich der übrigen Fehlerarten ergeben sich folgende Befunde:

Abbildung 4-15: Darstellung der Fehlerarten für die externe 'source monitoring' – Gruppe bei einer ersten Befragung nach einer (T1) bzw. nach sieben oder acht Wochen (NT2). Die Angaben beziehen sich auf beide Dia-Serien.



Auch beim Vergleich der verschiedenen Arten der Quellenverwechslungen (gesamt, innerhalb und zwischen den Dia-Serien) ergaben sich für keine dieser signifikante Veränderungen über die Zeit. Die Quellenverwechslungen gesamt und zwischen den Serien nehmen zu T1 gegenüber NT2 jeweils leicht ab, die Quellenverwechslungen innerhalb der Serien nehmen leicht zu.

Mit $T(20) = -2,714$ und $p = 0,007$ ist hingegen sogar ein hochsignifikanter Unterschied bezogen auf die unterschiedlichen Abstände einer ersten Befragung für das Vorkommen schematypischer Intrusionen zu verzeichnen. Nach acht Wochen kommen in der externalen 'source monitoring' - Gruppe weniger Intrusionen vor als nach einer Woche.

Konfabuliert wird in dieser Versuchsgruppe nach ca. acht Wochen allerdings mehr als nach einer Woche. Es lässt sich ein signifikanter Anstieg an Konfabulationen beobachten ($T(20) = 2,259$; $p = 0,018$).

Für die ESM-Gruppe lässt sich also festhalten, dass es hinsichtlich der korrekt berichteten Details insgesamt und im freien Bericht, der schematypischen Intrusionen sowie der Konfabulationen von Bedeutung ist, ob eine erste Befragung nach einer Woche oder erst nach sieben oder acht Wochen stattfindet.

Bezüglich der auf Nachfrage berichteten korrekten Details, der Quellenverwechslungen, der Gesamtfehlerzahl sowie der Rate der allgemeinen Erinnerungsfehler macht dies offensichtlich keinen so großen Unterschied.

Alle genauen Parameter und Ergebnisse sind in den *Tabellen 4-42* und *4-43* zu finden.

Tabelle 4-42: Mittelwerte und Standardabweichungen aller abhängigen Variablen für die externe 'source monitoring' – Gruppe zu T1 bzw. zu NT2. Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	44,00	8,19	32,36	13,55
korrekte Details im 'free recall'	18,55	7,56	10,64	5,75
korrekte Details 'auf Nachfrage'	25,45	5,99	21,82	9,01
Fehler gesamt	26,73	6,13	30,09	6,98
Quellenverwechslungen insgesamt	1,91	1,64	1,82	1,08
Quellenverwechslungen innerhalb	0,45	0,69	0,82	1,17
Quellenverwechslungen zwischen	1,45	1,29	1,00	0,78
schematypische Intrusionen	2,00	1,41	0,73	0,65
Konfabulationen	1,45	0,93	2,73	1,62
allgemeine Erinnerungsfehler	21,55	6,74	25,73	8,03

Tabelle 4-43: Ergebnisse des t –Tests für unabhängige Stichproben für alle Variablen für die externale ‘source monitoring’ Gruppe (T1 vs. NT2). Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	-2,438	20	0,012
korrekte Details im ‘free recall’	-2,763	20	0,006
korrekte Details ‘auf Nachfrage’	-1,115	20	0,139
Fehler gesamt	1,201	20	0,122
Quellenverwechslungen insgesamt	-0,154	20	0,440
Quellenverwechslungen innerhalb	0,890	16,189	0,194
Quellenverwechslungen zwischen	-1,000	20	0,165
schematypische Intrusionen	-2,714	20	0,007
Konfabulationen	2,259	20	0,018
allgemeine Erinnerungsfehler	1,323	20	0,101

Wie schon für die Veränderungen der Gedächtnisleistungen über die Zeit (4.1.3) erläutert, sollen auch hier im Folgenden die Ergebnisse für die beiden präsentierten Dia-Serien ‘PKW-PKW’ und ‘PKW-Fahrrad’ getrennt kurz dargestellt werden. Dazu wurden für jede Dia-Serie separat für alle abhängigen Variablen t-Tests für unabhängige Stichproben gerechnet. Es fanden sich folgende Ergebnisse:

Dia-Serie 1 (PKW-PKW):

Ebenso wie beim Vergleich der Gedächtnisleistungen über die Zeit (T1-T2) lässt sich auch hier vorweg festhalten, dass für keine der Variablen ein signifikanter Unterschied zwischen den Befragungszeitpunkten T1 und NT2 beobachtet werden kann. Die befragungszeitpunkt-bezogenen Veränderungen in dieser Gruppe sind für die Serie ‘PKW-PKW’ somit ebenfalls recht gering.

Jedoch ist tendenziell für folgende Variablen ein *leichter Anstieg in der Ausprägung des jeweiligen Mittelwerts* bezüglich T1 vs. NT2 zu verzeichnen:

- Fehler gesamt
- Quellenverwechslungen innerhalb der Dia-Serie
- Konfabulationen
- allgemeine Erinnerungsfehler

Ein *leichter Abfall im Mittelwert* fand sich hingegen bei folgenden Variablen:

- korrekte Details 'gesamt', im 'free recall' und 'auf Nachfrage'
- Quellenverwechslungen insgesamt und zwischen den Serien
- schematypische Intrusionen

Die genauen Mittelwerte, Standardabweichungen und auch die Ergebnisse der t-Tests für diese Serie sind nachfolgend tabellarisch aufgeführt und zusammengefasst (Tabellen 4-44 und 4-46).

Tabelle 4-44: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T1 und NT2 für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf die erste Dia-Serie 'PKW-PKW'.

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	23,18	6,27	20,27	9,23
korrekte Details im 'free recall'	8,82	3,87	6,36	3,72
korrekte Details 'auf Nachfrage'	14,36	6,99	13,91	6,02
Fehler gesamt	14,27	5,14	16,45	4,01
Quellenverwechslungen insgesamt	1,09	1,04	1,00	0,78
Quellenverwechslungen innerhalb	0,27	0,47	0,45	0,82
Quellenverwechslungen zwischen	0,82	0,87	0,55	0,69
schematypische Intrusionen	0,91	0,83	0,55	0,69
Konfabulationen	0,91	0,94	1,55	0,82
allgemeine Erinnerungsfehler	11,36	5,28	13,36	3,93

Dia-Serie 2 (PKW-Fahrrad):

Für diese zweite Dia-Serie ergibt sich vor allem für die Zahl der korrekt berichteten Details ein ganz anderes Bild.

Bezüglich dieser Serie macht es für die Versuchspersonen offenbar einen großen Unterschied in der Reproduktion korrekter Details, ob die erste Befragung nach einer oder nach ca. acht Wochen stattfindet. Zu NT2 sind die Mittelwerte für die korrekt berichteten Details (gesamt, im freien Bericht und auf Nachfrage) geringer als zu T1. Für die ersten beiden Variablen ergeben sich höchst- bzw. hochsignifikante Unterschiede ($T(20) = -3,652$; $p =$

0,001 bzw. $T(20) = -3,454$; $p = 0,002$), auf Nachfrage ein signifikanter Unterschied ($T(20) = -1,800$; $p = 0,044$).

Die Gesamtfehlerzahl steigt zu NT2 gegenüber T1 zwar leicht an, dieser Anstieg ist aber nicht signifikant.

Hinsichtlich der Quellenverwechslungen sind wiederum keinerlei signifikante Veränderungen erkennbar.

Dagegen weist der t-Test für das Vorkommen von schematypischen Intrusionen eine hochsignifikante Veränderung aus. Allerdings fallen diese zu NT2 geringer als zu T1.

Für die Variablen 'Konfabulationen' und 'allgemeine Erinnerungsfehler' lassen sich keine signifikanten Unterschiede zwischen den Befragungszeitpunkten konstatieren. Beide Mittelwerte sind aber bei einer ersten Befragung nach ca. acht Wochen höher als nach einer Woche.

Ausführlichere Informationen hierzu finden sich nachstehend in den *Tabellen 4-45 und 4-46*.

Tabelle 4-45: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T1 und NT2 für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf die zweite Dia-Serie 'PKW-Fahrrad'.

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	20,82	5,25	12,18	5,83
korrekte Details im 'free recall'	9,73	4,36	4,27	2,90
korrekte Details 'auf Nachfrage'	11,09	4,18	7,91	4,11
Fehler gesamt	12,45	3,21	13,64	4,59
Quellenverwechslungen insgesamt	0,82	0,87	0,82	1,33
Quellenverwechslungen innerhalb	0,18	0,41	0,36	0,92
Quellenverwechslungen zwischen	0,64	0,81	0,45	0,69
schematypische Intrusionen	1,09	0,83	0,18	0,41
Konfabulationen	0,55	0,69	1,18	1,33
allgemeine Erinnerungsfehler	10,18	2,96	11,45	3,59

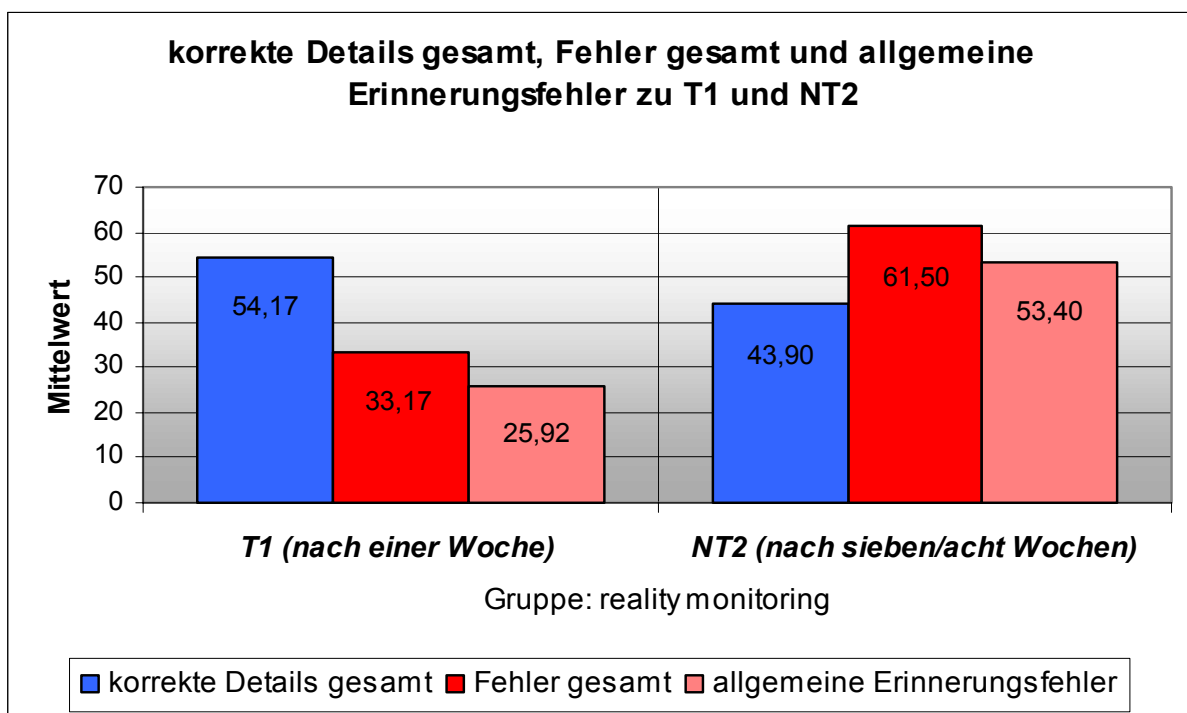
Tabelle 4-46: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Dia-Serien für die externe 'source-monitoring' – Gruppe zu T1 und NT2

Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	p
korrekte Details gesamt	-0,864	20	0,199	-3,652	20	0,001
korrekte Details im 'free recall'	-1,516	20	0,073	-3,454	20	0,002
korrekte Details 'auf Nachfrage'	-0,163	20	0,436	-1,800	20	0,044
Fehler gesamt	1,110	20	0,140	0,700	20	0,246
Quellenverwechslungen gesamt	-0,232	20	0,410	0,000	20	0,500
Quellenverwechslungen innerhalb	0,639	20	0,265	0,598	20	0,279
Quellenverwechslungen zwischen	-0,813	20	0,213	-0,568	20	0,288
schematypische Intrusionen	-1,118	20	0,139	-3,262	20	0,002
Konfabulationen	1,688	20	0,053	1,411	20	0,087
allgemeine Erinnerungsfehler	1,008	20	0,163	0,907	20	0,188

4.1.4.1.2 Ergebnisse für die 'reality monitoring' – Gruppe

Nach einer Woche (T1) wurden $n = 12$ Versuchspersonen interviewt, nach sieben bzw. acht Wochen (NT2) $n = 10$ Versuchspersonen. Es ergaben sich folgende Ergebnisse:

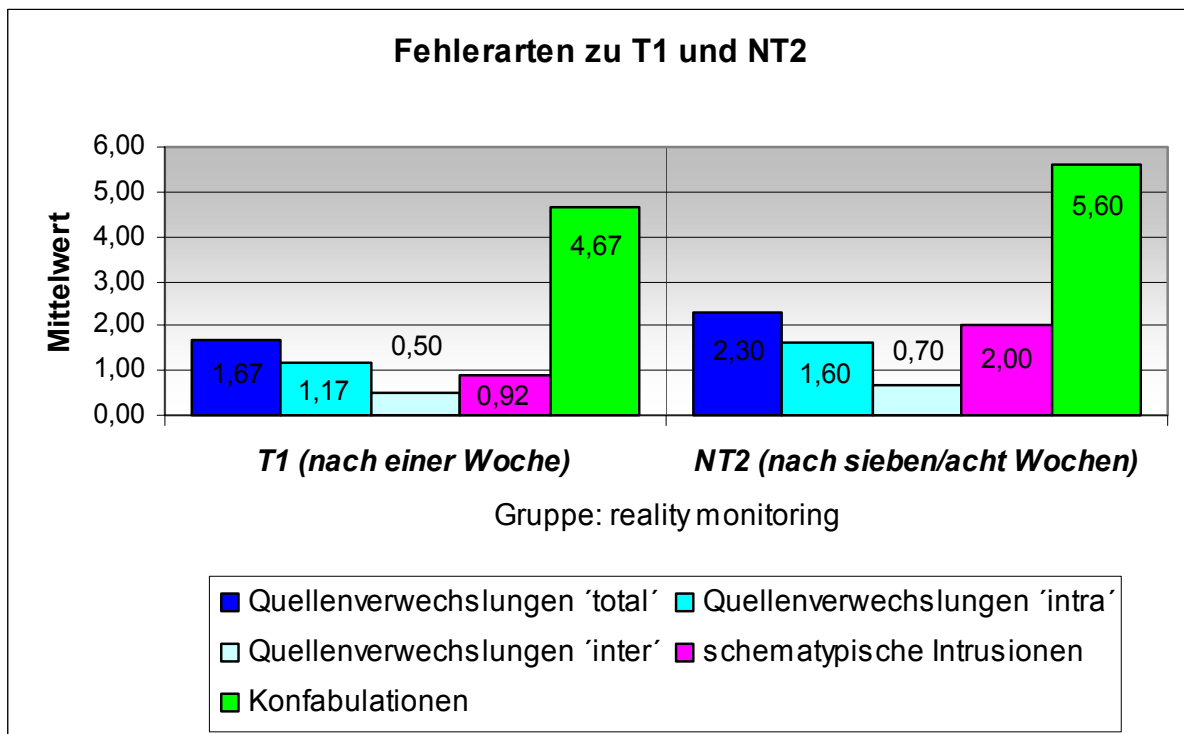
Abbildung 4-16: Darstellung der insgesamt berichteten korrekten Details, der insgesamt berichteten Fehler und der allgemeinen Erinnerungsfehler für die 'reality monitoring' – Gruppe bei einer ersten Befragung nach einer (T1) bzw. nach sieben oder acht Wochen (NT2). Die Angaben beziehen sich auf beide Unfall-Serien.



Vergleicht man die insgesamt korrekt berichteten Details ergibt sich erwartungsgemäß bei erstmaliger Befragung nach einer Woche gegenüber nach ca. acht Wochen ein Unterschied von mehr als 10 Details, die zu NT2 weniger berichtet werden, jedoch ist dieser Unterschied statistisch gesehen unbedeutend, weil nicht signifikant ($T(20) = -1,490$; $p = 0,076$). Auch im freien Bericht und auf Nachfrage sinken die Mittelwerte für T1 gegenüber NT2, aber auch diese bleiben nicht-signifikant.

Ganz anders dagegen gestalten sich die Ergebnisse für den Vergleich der Gesamtfehlerzahlen sowie der Raten der allgemeinen Erinnerungsfehler. Für beide Variablen ist ein sehr deutliches Ansteigen der Fehlerzahlen bei Befragung nach einer Woche gegenüber Befragung nach ca. acht Wochen zu beobachten. Dies äußert sich in beiden Fällen in einem höchstsignifikanten Ergebnis des t-Tests ($T(20) = 9,842$; $p < 0,001$ und $T(20) = 8,947$; $p < 0,001$).

Abbildung 4-17: Darstellung der Fehlerarten für die 'reality monitoring' – Gruppe bei einer ersten Befragung nach einer (T1) bzw. nach sieben oder acht Wochen (NT2). Die Angaben beziehen sich auf beide Unfall-Serien.



Bei Betrachtung der verschiedenen Arten von Quellenverwechslungen (gesamt, innerhalb und zwischen den Serien) ergibt sich für keine der genannten eine signifikante Veränderung von T1 gegenüber NT2. Für alle drei kann aber festgestellt werden, dass sich die Mittelwerte jeweils erhöhen.

Ebenfalls ein Anstieg des Mittelwerts ist hinsichtlich der schematypischen Intrusionen zu verzeichnen, er verdoppelt sich in etwa bei Befragung nach einer Woche gegenüber einer ersten Befragung nach acht Wochen. Dieser Unterschied ist mit $T(13,571) = 2,156$ und $p = 0,025$ signifikant.

Zwar zeigen sich über beide Serien hinweg nach acht Wochen auch mehr Konfabulationen als nach einer Woche, jedoch wird das angenommene Signifikanzniveau von 0,05 verfehlt. Der Unterschied erweist sich als nicht-signifikant ($T(20) = 0,610$; $p = 0,274$).

Die genauen und vollständigen Werte und Ergebnisse lassen sich in den nachstehenden Tabellen 4-47 und 4-48 nachlesen:

Tabelle 4-47: Mittelwerte und Standardabweichungen aller abhängigen Variablen für die 'reality monitoring' – Gruppe zu T1 bzw. zu NT2. Die Angaben beziehen sich auf beide Unfall-Serien.

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	54,17	17,49	43,90	14,21
korrekte Details im 'free recall'	18,42	9,69	15,50	6,13
korrekte Details 'auf Nachfrage'	35,75	13,90	28,60	10,11
Fehler gesamt	33,17	5,73	61,50	7,76
Quellenverwechslungen insgesamt	1,67	1,44	2,30	2,06
Quellenverwechslungen innerhalb	1,17	1,03	1,60	1,27
Quellenverwechslungen zwischen	0,50	0,80	0,70	1,06
Schematypische Intrusionen	0,92	0,79	2,00	1,41
Konfabulationen	4,67	2,81	5,60	4,33
allgemeine Erinnerungsfehler	25,92	5,25	53,40	8,98

Tabelle 4-48: Ergebnisse des t –Tests für unabhängige Stichproben für alle Variablen für die 'reality monitoring' – Gruppe (T1 vs. NT2). Die Angaben beziehen sich auf beide Unfall-Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	-1,490	20	0,076
korrekte Details im 'free recall'	-0,823	20	0,210
korrekte Details 'auf Nachfrage'	-1,353	20	0,096
Fehler gesamt	9,842	20	< 0,001
Quellenverwechslungen insgesamt	0,849	20	0,203
Quellenverwechslungen innerhalb	0,886	20	0,193
Quellenverwechslungen zwischen	0,505	20	0,310
Schematypische Intrusionen	2,156	13,571	0,025
Konfabulationen	0,610	20	0,274
allgemeine Erinnerungsfehler	8,947	20	< 0,001

Analog zu den vorangegangenen Ausführungen sollen im Folgenden auch für die RM-Gruppe die Ergebnisse der t-Tests für die beiden Unfall-Serien getrennt voneinander betrachtet und dargestellt werden.

Dia-Serie 1 (PKW-PKW):

Für diese Dia-Serie ergeben sich hinsichtlich der korrekt berichteten Details – sei es insgesamt, im freien Bericht als auch auf Nachfrage - keinerlei statistisch bedeutsamen, also signifikante Veränderungen zu den unterschiedlichen Befragungszeitpunkten T1 und NT2. Die Unterschiede in den Mittelwerten bewegen sich nach oben und nach unten in recht kleinen Ausmaßen. Es scheint hier also diesbezüglich keine Rolle zu spielen, ob eine erste Befragung schon nach einer Woche oder erst nach ca. acht Wochen stattfindet.

Hingegen verändert sich die Zahl der insgesamt begangenen Fehler zu T1 gegenüber NT2 recht stark. Die Versuchspersonen machen demzufolge nach ca. acht Wochen höchstsignifikant mehr Fehler als nach einer Woche ($T(20) = 4,968$; $p < 0,001$).

Des Weiteren sind für alle Arten der Quellenverwechslungen auch hier wiederum keine signifikanten Unterschiede zwischen T1 und NT2 festzustellen bzw. lassen sich aufgrund des Mittelwerts 0 nicht berechnen. Bei den Quellenverwechslungen insgesamt und innerhalb der Dia-Serie sind nur sehr minimale Veränderungen vorhanden, bei denen zwischen den Serien überhaupt keine.

Weiterhin kommen bei den Versuchspersonen der ESM-Gruppe nach acht Wochen erwartungsgemäß signifikant mehr schematypische Intrusionen zustande als nach einer Woche ($T(20) = 1,992$; $p = 0,030$).

Der Mittelwert für die Variable 'Konfabulationen' ist bei erstmaliger Befragung zu NT2 nach acht Wochen für diese Dia-Serie zwar höher als zu T1 nach einer Woche, allerdings ergibt sich daraus kein signifikantes Ergebnis.

Wie die Gesamtfehlerzahl ist auch die Rate der allgemeinen Erinnerungsfehler mit knapp sieben Fehlern mehr zu NT2 sehr viel höher als zu T1, was sich statistisch und hypothesenkonform in einem höchstsignifikanten Ergebnis niederschlägt ($T(20) = 4,231$; $p < 0,001$).

Die nachfolgenden *Tabellen 4-49* und *4-51* verdeutlichen und ergänzen oben stehende Ausführungen zur Dia-Serie 'PKW-PKW'.

Tabelle 4-49: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T1 und NT2 für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 1 (PKW-PKW).

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	35,83	9,58	34,10	12,41
korrekte Details im 'free recall'	11,17	5,10	11,80	6,18
korrekte Details 'auf Nachfrage'	24,67	8,47	22,30	8,06
Fehler gesamt	15,67	3,06	24,40	5,10
Quellenverwechslungen insgesamt	0,67	0,99	0,90	0,99
Quellenverwechslungen innerhalb	0,67	0,99	0,90	0,99
Quellenverwechslungen zwischen	0,00	0,00	0,00	0,00
Schematypische Intrusionen	0,50	0,67	1,10	0,74
Konfabulationen	1,75	1,49	2,70	1,89
allgemeine Erinnerungsfehler	12,75	3,19	19,70	4,50

Serie 2 - selbst ausgedachter Unfall – analog zu Dia-Serie 'PKW-Fahrrad':

Mussten sich Versuchspersonen die zweite Unfallserie selbst ausdenken, scheinen sie bei einer ersten Befragung nach einer Woche deutlich weniger Probleme zu haben, korrekte Details zu erinnern und zu berichten als bei dieser ersten Befragung nach sieben oder acht Wochen. Dies gilt dabei sowohl für die Gesamtzahl der korrekten Details, als auch für die, die im free recall und auch auf Nachfrage berichtet wurden. Die Mittelwerte der drei Variablen sind zu NT2 jeweils deutlich niedriger als zu T1. Es ergeben sich somit hoch- bzw. signifikante Ergebnisse – für die insgesamt korrekt berichteten Details: $T(17,254) = -2,580$; $p = 0,010$, im free recall: $T(20) = -2,002$; $p = 0,030$, auf Nachfrage: $T(16,253) = -1,828$; $p = 0,043$.

Gleiches gilt für den Vergleich der Gesamtfehlerzahlen für diese selbst erfundene Unfallserie. Auch hier ist die Fehlerrate nach acht Wochen sehr viel höher (im Schnitt über 20 Fehler mehr) als nach einer Woche. Der t-Test weist demnach mit $T(12,048) = 7,423$ und $p < 0,001$ hierfür ein höchstsignifikantes Ergebnis aus.

Für die Quellenverwechslungen (gesamt, innerhalb und zwischen den Serien) ergibt sich jeweils keine signifikante Veränderung beim Vergleich der Befragungszeitpunkte T1 und NT2. Jedoch sind alle Mittelwerte bei erster Befragung nach acht Wochen höher als nach einer Woche.

Ein ähnliches Bild zeigt sich hinsichtlich der berichteten schematypischen Intrusionen und auch Konfabulationen. Hier lassen sich jeweils ein leichter Anstieg der Mittelwerte bei T1 gegenüber NT2 feststellen, welche aber nicht signifikant sind.

Bei Betrachtung der allgemeinen Erinnerungsfehler wiederum, beobachten wir wie bei der Gesamtfehlerzahl einen deutlichen, auch statistisch sogar höchstsignifikanten Anstieg der Fehlerzahl im Vergleich der beiden Befragungszeitpunkte ($T(11,106) = 6,557$; $p < 0,001$). Nach acht Wochen liegt die Rate der allgemeinen Erinnerungsfehler im Mittel um gut 20 höher als bei einer ersten Befragung nach einer Woche.

Weitere, genauere Werte dazu finden sich wie gehabt in den nachstehenden *Tabellen 4-50* und *4-51*.

Tabelle 4-50: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T1 und NT2 für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf die Serie 2 – analog Dia-Serie 'PKW-Fahrrad'.

Variable	T1		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	18,33	9,91	9,80	5,25
korrekte Details im 'free recall'	7,25	5,14	3,50	3,21
korrekte Details 'auf Nachfrage'	11,08	8,05	6,30	3,80
Fehler gesamt	17,50	3,85	39,10	8,50
Quellenverwechslungen insgesamt	1,00	0,95	1,40	1,65
Quellenverwechslungen innerhalb	0,50	0,67	0,70	0,82
Quellenverwechslungen zwischen	0,50	0,80	0,70	1,06
Schematypische Intrusionen	0,42	0,52	0,90	0,99
Konfabulationen	2,92	2,11	3,10	2,33
allgemeine Erinnerungsfehler	13,17	3,51	33,70	9,37

Tabelle 4-51: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Serien für die 'reality monitoring' – Gruppe zu T1 und NT2

Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	p
korrekte Details gesamt	-0,370	20	0,358	-2,580	17,254	0,010
korrekte Details im 'free recall'	-0,264	20	0,398	-2,002	20	0,030
korrekte Details 'auf Nachfrage'	-0,667	20	0,256	-1,828	16,253	0,043
Fehler gesamt	4,968	20	< 0,001	7,423	12,048	< 0,001
Quellenverwechslungen gesamt	0,551	20	0,294	0,712	20	0,242
Quellenverwechslungen innerhalb	0,551	20	0,294	0,627	20	0,269
Quellenverwechslungen zwischen	<i>n. berechenbar MWe = 0</i>			0,505	20	0,310
Schematypische Intrusionen	1,992	20	0,030	1,469	20	0,079
Konfabulationen	1,322	20	0,101	0,194	20	0,424
allgemeine Erinnerungsfehler	4,231	20	< 0,001	6,557	11,106	< 0,001

4.1.4.2 Befragungszeitpunkt T2 (nach sieben/acht Wochen bei vorausgehender Befragung zu T1) vs. Befragungszeitpunkt NT2 (nach sieben bzw. acht Wochen ohne vorausgehende Befragung)

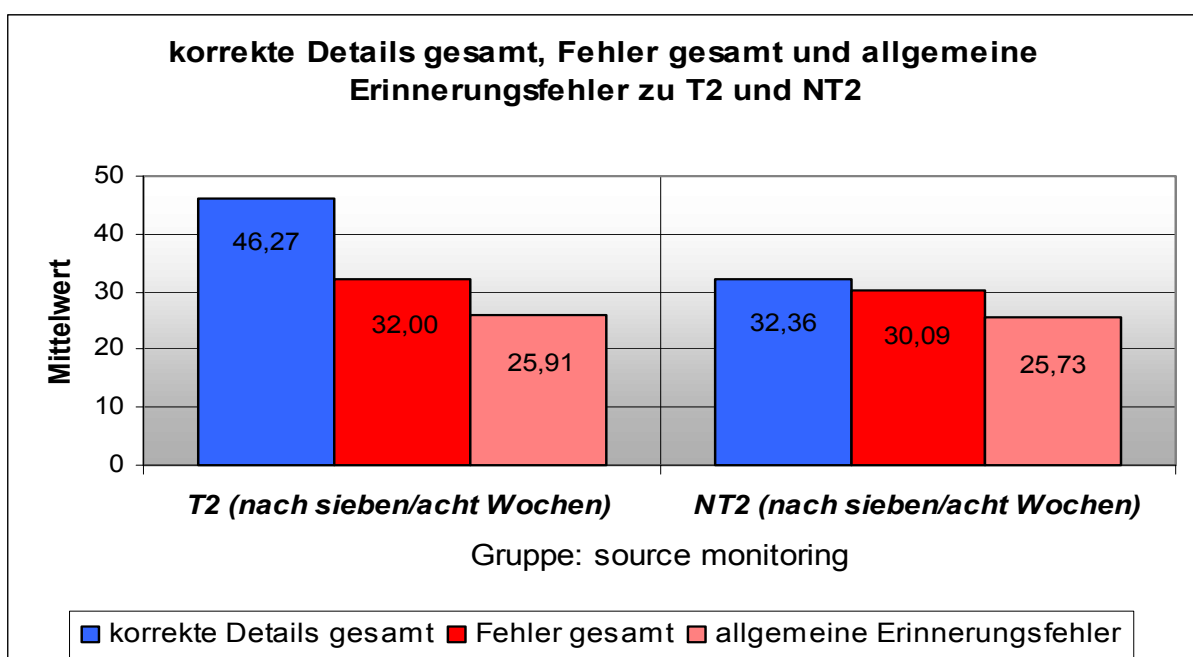
Ebenfalls interessant ist der Vergleich der beiden Befragungszeitpunkte T2 und NT2, also die Frage, ob und in welchen Aspekten der Gedächtnisleistung sich Unterschiede zeigen zwischen den Versuchspersonen, die nach sieben oder acht Wochen ein zweites Mal befragt wurden (T2) und denjenigen, die zum selben Zeitpunkt das erste Mal interviewt wurden (NT2).

Dazu wurden wiederum getrennt für beide Versuchsgruppen (externe 'source monitoring' und 'reality monitoring') t-Tests für unabhängige Stichproben durchgeführt. Die Ergebnisse dessen werden im Folgenden für beide Gruppen dargestellt.

4.1.4.2.1 Ergebnisse für die externe 'source monitoring' – Gruppe

Zu T2 wurden $n = 11$ Versuchspersonen befragt, zu NT2 ebenfalls $n = 11$.

Abbildung 4-18: Darstellung der insgesamt berichteten korrekten Details, der insgesamt berichteten Fehler und der allgemeinen Erinnerungsfehler für die externe 'source monitoring' – Gruppe bei Befragung nach sieben/acht Wochen bei vorangegangener Befragung (T2) bzw. nach sieben/acht Wochen ohne vorangegangene Befragung (NT2). Die Angaben beziehen sich auf beide Dia-Serien.



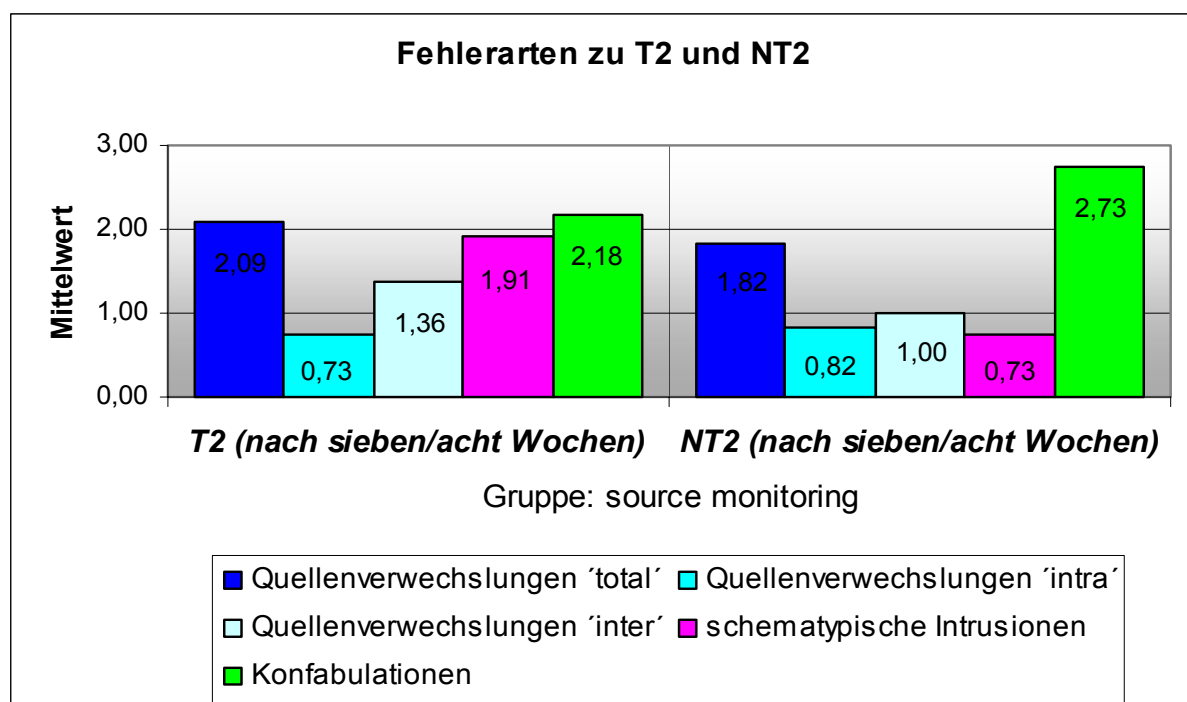
Bezogen auf beide Dia-Serien berichten die Versuchspersonen der ESM-Gruppe bei zweiter Befragung nach ca. acht Wochen hochsignifikant mehr korrekte Details gesamt als bei einer ersten Befragung nach acht Wochen ($T(20) = -2,538$; $p = 0,010$). Ein ähnlicher, noch größerer gleichgerichteter Effekt zeigt sich hier bei den korrekt berichteten Details im free recall – der Unterschied ist mit $T(20) = -3,983$ und $p < 0,001$ sogar höchstsignifikant. Kein signifikantes Ergebnis findet sich bezüglich der korrekten Details auf Nachfrage, jedoch sind die Erinnerungsleistungen zu NT2 auch hier schwächer als zu T2.

Hinsichtlich der Gesamtfehlerzahl sowie der Zahl der allgemeinen Erinnerungsfehler scheint ein analoger Zusammenhang nicht zu gelten. Es macht hier offensichtlich für beide Variablen keinen (signifikanten) Unterschied, ob nach ca. acht Wochen zum ersten Mal befragt wird oder ob vorher schon eine Befragung stattgefunden hat ($T(20) = -0,685$; $p = 0,251$ bzw. $T(20) = -0,063$; $p = 0,475$). Die Fehlerraten sind beide Male zu NT2 sogar etwas niedriger als zu T2.

Genauere Informationen zu diesen einzelnen Teilergebnissen finden sich in den Tabellen 4-52 und 4-53.

Für die übrigen Fehlerarten ergaben sich folgende Befunde:

Abbildung 4-19: Darstellung der Fehlerarten für die externe 'source monitoring' – Gruppe bei Befragung nach sieben/acht Wochen bei vorangegangener Befragung (T2) bzw. nach sieben/acht Wochen ohne vorangegangene Befragung (NT2). Die Angaben beziehen sich auf beide Dia-Serien.



Untersucht man für beide Dia-Serien ob und welche Quellenverwechslungen vorkommen, so stellt man fest, dass sich abermals zwischen Befragungszeitpunkten keine statistisch relevanten Veränderungen finden lassen. Die t-Tests weisen für alle Arten der Quellenverwechslungen (gesamt, innerhalb und zwischen den Dia-Serien) keine signifikanten Ergebnisse aus. Im Weiteren finden sich bei einer ersten und einzigen Befragung nach ca. acht Wochen deutlich weniger schematypische Intrusionen als bei einer zweiten zum selben Zeitpunkt. Mit $T(20) = -3,191$ und $p = 0,003$ erweist sich dieser Unterschied als hochsignifikant.

Konfabulationen lassen sich nach sieben bzw. acht Wochen zu NT2 zwar etwas mehr beobachten als zu T2, allerdings ergibt sich im t-Test kein signifikantes Ergebnis.

Die Tabellen 4-52 und 4-53 enthalten sämtliche Kennwerte der Variablen sowie die Ergebnisse der jeweiligen t-Tests.

Tabelle 4-52: Mittelwerte und Standardabweichungen aller abhängigen Variablen für die externe 'source monitoring' – Gruppe zu T2 bzw. zu NT2. Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	46,27	12,11	32,36	13,55
korrekte Details im 'free recall'	22,27	7,80	10,64	5,75
korrekte Details 'auf Nachfrage'	24,00	10,19	21,82	9,01
Fehler gesamt	32,00	6,07	30,09	6,98
Quellenverwechslungen insgesamt	2,09	1,70	1,82	1,08
Quellenverwechslungen innerhalb	0,73	1,27	0,82	1,17
Quellenverwechslungen zwischen	1,36	1,29	1,00	0,78
Schematypische Intrusionen	1,91	1,04	0,73	0,65
Konfabulationen	2,18	2,27	2,73	1,62
allgemeine Erinnerungsfehler	25,91	5,21	25,73	8,03

Tabelle 4-53: Ergebnisse des t –Tests für unabhängige Stichproben für alle Variablen für die externale ‘source monitoring’ Gruppe (T2 vs. NT2). Die Angaben beziehen sich auf beide Dia-Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	-2,538	20	0,010
korrekte Details im ‘free recall’	-3,983	20	< 0,001
korrekte Details ‘auf Nachfrage’	-0,532	20	0,301
Fehler gesamt	-0,685	20	0,251
Quellenverwechslungen insgesamt	-0,449	20	0,329
Quellenverwechslungen innerhalb	0,175	20	0,432
Quellenverwechslungen zwischen	-0,803	20	0,216
Schematypische Intrusionen	-3,191	20	0,003
Konfabulationen	0,649	20	0,262
allgemeine Erinnerungsfehler	-0,063	20	0,475

Getrennt nach Dia-Serien lassen sich die folgenden Befunde und Ergebnisse der t-Tests für unabhängige Stichproben festhalten und beschreiben:

Dia-Serie 1 (PKW-PKW):

Insgesamt lässt sich für diese Dia-Serie vorwegnehmen, dass sich außer für die korrekten Details im freien Bericht ($T(20) = -3,322$ und $p = 0,003$) für keine der Variablen ein signifikantes Testergebnis bezüglich einer ersten gegenüber einer zweiten Befragung nach ca. acht Wochen herausstellt. Die Unterschiede erweisen sich insgesamt als doch gering.

Bei einmaliger und erster Befragung zu NT2 werden zwar im Durchschnitt knapp fünf Details insgesamt weniger erinnert bzw. berichtet, jedoch verfehlt der t-Test das Signifikanzniveau. Des Weiteren können außerdem für folgende Variablen zu T2 höhere Mittelwerte beobachtet werden als zu NT2:

- korrekte Details im free recall
- Quellenverwechslungen ‘gesamt’, ‘innerhalb’ und ‘zwischen’ den Dia-Serien
- schematypische Intrusionen

Zu NT2 höhere Mittelwerte finden sich für:

- korrekte Details auf Nachfrage
- Fehler gesamt und allgemeine Erinnerungsfehler
- Konfabulationen

Vollständige und ausführlichere Informationen sind diesbezüglich tabellarisch in den Tabellen 4-54 und 4-55 dargestellt.

Tabelle 4-54: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T2 und NT2 für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 1 (PKW-PKW).

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	25,18	6,59	20,27	9,23
korrekte Details im 'free recall'	11,64	3,72	6,36	3,72
korrekte Details 'auf Nachfrage'	13,55	6,55	13,91	6,02
Fehler gesamt	15,64	3,59	16,45	4,01
Quellenverwechslungen insgesamt	1,27	1,49	1,00	0,78
Quellenverwechslungen innerhalb	0,55	1,21	0,45	0,82
Quellenverwechslungen zwischen	0,73	0,91	0,55	0,69
Schematypische Intrusionen	0,73	0,79	0,55	0,69
Konfabulationen	1,27	1,68	1,55	0,82
allgemeine Erinnerungsfehler	12,36	3,56	13,36	3,93

Dia-Serie 2 (PKW-Fahrrad):

Etwas anders gestalten sich die Ergebnisse für die Dia-Serie 'PKW-Fahrrad'.

Hier ergibt sich beim Vergleich der insgesamt korrekt berichteten Details zwischen T2 und NT2 eine hochsignifikante Veränderung ($T(20) = -3,047$; $p = 0,003$). Bei der einmaligen Befragung nach ca. acht Wochen berichten die Versuchspersonen wesentlich weniger korrekte Details als bei einer zweiten. Gleiches gilt für die korrekten Details, die im freien Bericht erinnert werden – auch hier ergibt sich mit $T(14,783) = -3,277$ und $p = 0,003$ ein hochsignifikantes Ergebnis. Eine ähnliche Tendenz weisen auch die auf Nachfrage berichteten korrekten Details auf, jedoch ergibt sich hier kein statistisch signifikantes Ergebnis.

Analog zur Dia-Serie 'PKW-PKW' lassen sich hinsichtlich der Gesamtfehlerzahl und der Zahl der allgemeinen Erinnerungsfehler keine signifikanten Unterschiede konstatieren, wobei die Versuchspersonen hier bei einer ersten Befragung nach acht Wochen (NT2) jeweils weniger Fehler insgesamt und auch weniger allgemeine Erinnerungsfehler begehen.

Bezüglich der Quellenverwechslungen können im Ergebnis weiterhin auch keine statistisch bedeutsamen Veränderungen festgehalten werden.

Gleiches gilt für das Vorkommen von Konfabulationen zu den einzelnen Zeitpunkten.

Ein allerdings hochsignifikanter Unterschied lässt sich für die schematypischen Intrusionen beschreiben. Nach acht Wochen machen die Versuchspersonen, die dann bereits ein zweites mal befragt wurden mehr Fehler diesbezüglich als diejenigen, die ein erstes Mal interviewt wurden ($T(14,098) = -3,444$; $p = 0,002$).

Eine detailliertere tabellarische Zusammenfassung findet sich nachstehend (Tabellen 4-55 und 4-56).

Tabelle 4-55: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T2 und NT2 für die externe 'source monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 2 (PKW-Fahrrad).

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	21,09	7,75	12,18	5,83
korrekte Details im 'free recall'	10,64	5,75	4,27	2,90
korrekte Details 'auf Nachfrage'	10,45	6,74	7,91	4,11
Fehler gesamt	16,36	4,50	13,64	4,59
Quellenverwechslungen insgesamt	0,82	0,98	0,82	1,33
Quellenverwechslungen innerhalb	0,18	0,60	0,36	0,92
Quellenverwechslungen zwischen	0,64	0,67	0,45	0,69
Schematypische Intrusionen	1,18	0,87	0,18	0,41
Konfabulationen	0,91	0,94	1,18	1,33
allgemeine Erinnerungsfehler	13,55	4,63	11,45	3,59

Tabelle 4-56: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Dia-Serien für die externe 'source-monitoring' – Gruppe zu T2 und NT2

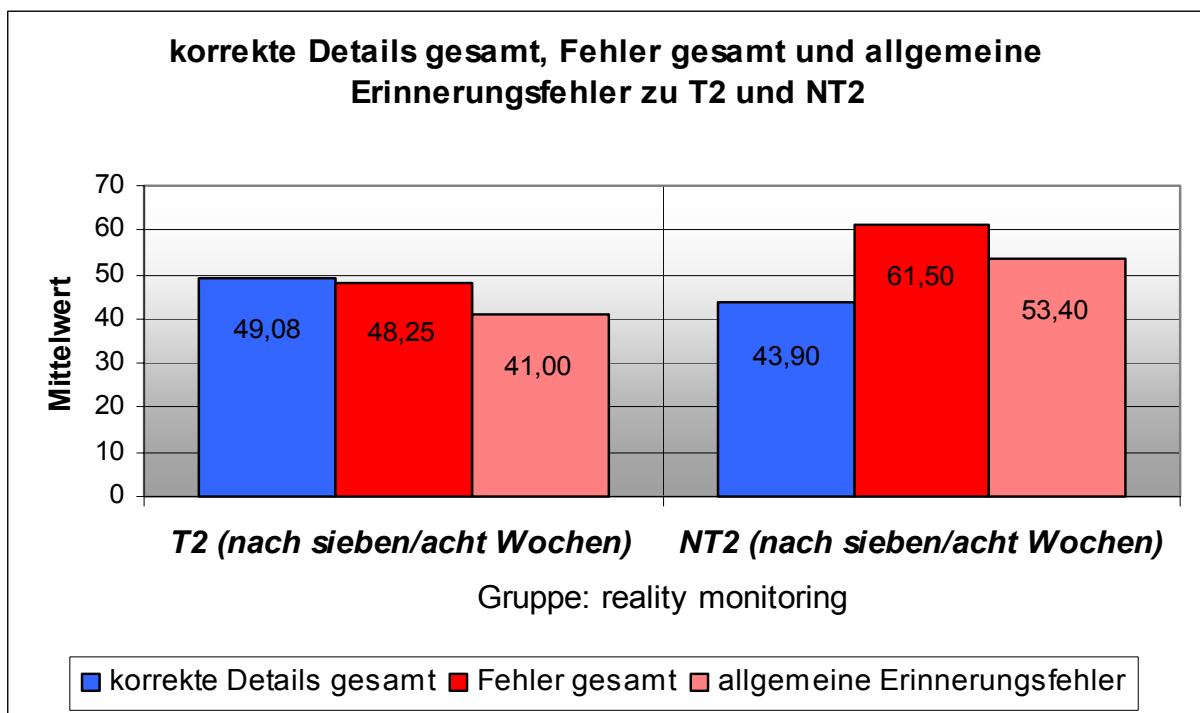
Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	p
korrekte Details gesamt	-1,436	20	0,084	-3,047	20	0,003
korrekte Details im 'free recall'	-3,322	20	0,002	-3,277	14,783	0,003
korrekte Details 'auf Nachfrage'	0,136	20	0,447	-1,069	20	0,149
Fehler gesamt	0,505	20	0,301	-1,407	20	0,088
Quellenverwechslungen gesamt	-0,539	15,041	0,299	0,000	20	0,500
Quellenverwechslungen innerhalb	-0,206	20	0,420	0,546	20	0,296
Quellenverwechslungen zwischen	-0,531	20	0,301	-0,626	20	0,269
Schematypische Intrusionen	-0,577	20	0,285	-3,444	14,098	0,002
Konfabulationen	0,484	20	0,317	0,555	20	0,293
allgemeine Erinnerungsfehler	0,626	20	0,270	-1,183	20	0,126

4.1.4.2.2 Ergebnisse für die 'reality monitoring' – Gruppe

Die RM-Gruppe umfasste zum Befragungszeitpunkt T2 $n = 12$ Versuchspersonen, zum Befragungszeitpunkt NT2 $n = 10$ Versuchspersonen.

Nach Auswertung und Durchführung der t-Tests stellte sich folgendes Bild dar:

Abbildung 4-20: Darstellung der insgesamt berichteten korrekten Details, der insgesamt berichteten Fehler und der allgemeinen Erinnerungsfehler für die 'reality monitoring' – Gruppe bei Befragung nach sieben/acht Wochen bei vorangegangener Befragung (T2) bzw. nach sieben/acht Wochen ohne vorangegangene Befragung (NT2). Die Angaben beziehen sich auf beide Unfall-Serien.



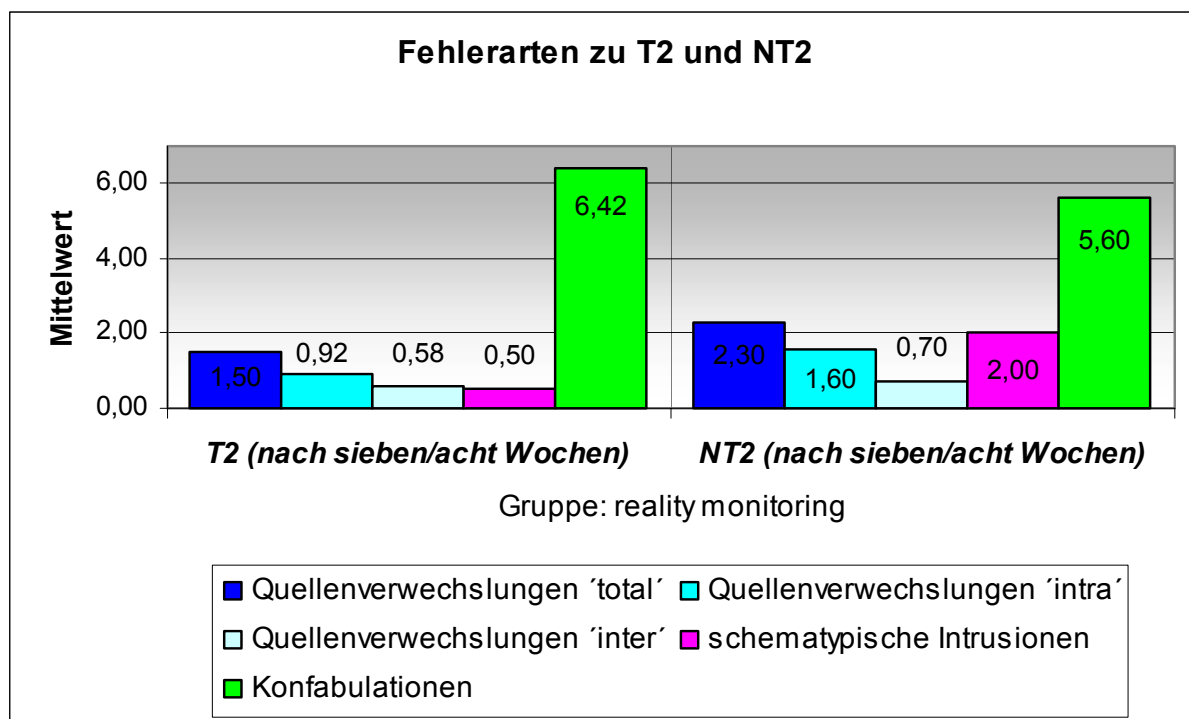
Vergleicht man die Gesamtzahl der korrekt berichteten Details für die beiden Zeitpunkte T2 und NT2 miteinander, so kann kein signifikanter Unterschied festgestellt werden ($T(20) = -0,787$; $p = 0,221$). Gleiches gilt bei differenzierter Betrachtung für die korrekt berichteten Details sowohl im freien Bericht als auch auf Nachfrage hin. Zwar erinnern die Versuchspersonen hypothesengemäß bei einer zweiten Befragung nach ca. acht Wochen jeweils mehr Details (gesamt, im free recall und auf Nachfrage) als diejenigen bei einer ersten Befragung nach demselben Zeitraum, allerdings gestalten sich eben diese Unterschiede nie signifikant. Dahingegen scheinen hier ein weiteres Mal die Fehlerzahlen eine entscheidende Bedeutung zu haben. Hinsichtlich der Gesamtfehlerzahl ergibt sich - wie in der Fragestellung angenom-

men - zu NT2 erneut ein wesentlich höherer Wert als zu T2, was sich statistisch in einem höchstsignifikanten Unterschied äußert ($T(20) = 4,653$; $p < 0,001$).

Ebenso unterlaufen den Versuchspersonen, die nach sieben oder acht Wochen das erste Mal befragt wurden, sehr viel mehr allgemeine Erinnerungsfehler als diejenigen, die dann schon ein zweites Mal befragt wurden. Dies verdeutlicht statistisch das höchstsignifikante Ergebnis des t-Tests ($T(20) = 3,738$; $p < 0,001$).

Für Quellenverwechslungen, Intrusionen und Konfabulationen ergaben sich nachfolgende Befunde:

Abbildung 4-21: Darstellung der Fehlerarten für die 'reality monitoring' – Gruppe bei Befragung nach sieben/acht Wochen bei vorangegangener Befragung (T2) bzw. nach sieben/acht Wochen ohne vorangegangene Befragung (NT2). Die Angaben beziehen sich auf beide Unfall-Serien.



Lediglich für den Vergleich von genannten schematypischen Intrusionen ergibt sich hier eine (hoch-)signifikante Veränderung ($T(20) = 3,015$; $p = 0,004$) in Form eines vierfachen Anstiegs bei einmaliger Befragung (NT2) nach acht Wochen gegenüber einer zweiten (T2).

Quellenverwechslungen aller Art sowie die Zahl der geäußerten Konfabulationen unterscheiden sich jedoch nicht wesentlich oder statistisch bedeutsam zu T2 und NT2. Dabei kann tendenziell bemerkt werden, dass die einzelnen Werte der Quellenverwechslungen (gesamt, innerhalb und zwischen den Serien) zu NT2 geringer sind als zu T2. Das Gegenteil gilt hinsichtlich der Konfabulationen.

Die genauen Mittelwerte, Standardabweichungen und Signifikanzergebnisse für alle abhängigen Variablen sind nachstehend tabellarisch aufgeführt.

Tabelle 4-57: Mittelwerte und Standardabweichungen aller abhängigen Variablen für die 'reality monitoring' – Gruppe zu T2 bzw. zu NT2. Die Angaben beziehen sich auf beide Unfall - Serien.

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	49,08	16,28	43,90	14,21
korrekte Details im 'free recall'	19,50	8,97	15,50	6,13
korrekte Details 'auf Nachfrage'	29,58	9,98	28,60	10,11
Fehler gesamt	48,25	5,58	61,50	7,76
Quellenverwechslungen insgesamt	1,50	1,31	2,30	2,06
Quellenverwechslungen innerhalb	0,92	1,44	1,60	1,27
Quellenverwechslungen zwischen	0,58	0,79	0,70	1,06
Schematypische Intrusionen	0,50	0,91	2,00	1,41
Konfabulationen	6,42	3,26	5,60	4,33
allgemeine Erinnerungsfehler	41,00	6,56	53,40	8,98

Tabelle 4-58: Ergebnisse des t –Tests für unabhängige Stichproben für alle Variablen für die 'reality monitoring' - Gruppe (T2 vs. NT2). Die Angaben beziehen sich auf beide Unfall – Serien.

Variable	T	df	Signifikanz
korrekte Details gesamt	-0,787	20	0,221
korrekte Details im 'free recall'	-1,194	20	0,123
korrekte Details 'auf Nachfrage'	-0,229	20	0,411
Fehler gesamt	4,653	20	< 0,001
Quellenverwechslungen insgesamt	1,106	20	0,141
Quellenverwechslungen innerhalb	1,168	20	0,128
Quellenverwechslungen zwischen	0,295	20	0,386
Schematypische Intrusionen	3,015	20	0,004
Konfabulationen	-0,505	20	0,310
allgemeine Erinnerungsfehler	3,738	20	< 0,001

Wie bereits in den voran stehenden Ausführungen sollen im Folgenden die Ergebnisse für – in diesem Fall die RM-Gruppe – die beiden Unfall-Serien (Dia-Serie 1 (PKW-PKW) sowie die Serie 2 (selbst erfundene Unfallserie – analog der Dia-Serie 'PKW-Fahrrad')) getrennt voneinander betrachtet werden.

Hierzu ergibt sich im Einzelnen Folgendes:

Dia-Serie 1 (PKW-PKW):

Anhand der durchgeführten t-Tests ließen sich für die korrekt berichteten Details (gesamt, im freien Bericht und auf Nachfrage) jeweils keine signifikanten Ergebnisse beim Vergleich der Befragungszeitpunkte T2 und NT2 ermitteln. Tendenziell waren die Mittelwerte zu NT2 jeweils geringer als zu T2.

Für alle Fehlervariablen dagegen konnten für diese Dia-Serie bei einer ersten Befragung nach ca. acht Wochen (NT2) zwar jeweils höhere Mittelwerte festgestellt werden als bei einer zweiten Befragung ebenfalls nach ca. acht Wochen (T2), jedoch erweisen sich dabei nur zwei der Unterschiede als signifikant bzw. hochsignifikant:

Was die Gesamtfehlerzahl betrifft, ergibt sich hier ein signifikanter Unterschied ($T(20) = 2,000$; $p = 0,030$).

Hochsignifikant wird mit $T(20) = 3,320$ und $p = 0,002$ das Ergebnis für die Rate der schematischen Intrusionen.

Quellenverwechslungen, Konfabulationen und allgemeine Erinnerungsfehler unterscheiden sich zu den genannten Zeitpunkten nicht besonders stark oder gar nicht (nicht-signifikant).

Im Anschluss die konkreten Stichprobenkennwerte und t-Test-Ergebnisse (*Tabellen 4-59 und 4-61*).

Tabelle 4-59: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T2 und NT2 für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf Dia-Serie 1 (PKW-PKW).

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	36,25	10,81	34,10	12,41
korrekte Details im 'free recall'	13,75	7,06	11,80	6,18
korrekte Details 'auf Nachfrage'	22,50	6,87	22,30	8,06
Fehler gesamt	20,33	4,44	24,40	5,10
Quellenverwechslungen insgesamt	0,42	0,67	0,90	0,99
Quellenverwechslungen innerhalb	0,42	0,67	0,90	0,99
Quellenverwechslungen zwischen	0,00	0,00	0,00	0,00
Schematypische Intrusionen	0,25	0,45	1,10	0,74
Konfabulationen	2,33	1,78	2,70	1,89
allgemeine Erinnerungsfehler	17,33	4,23	19,70	4,50

Serie 2 (selbst ausgedachte Unfall-Serie – analog zu Dia-Serie 2 'PKW-Fahrrad'):

Bei Betrachtung der korrekt berichteten Details ergibt sich hier dasselbe Endergebnis wie für die Dia-Serie. Es findet sich weder in der Gesamtzahl, noch im freien Bericht oder auf Nachfrage ein statistisch bedeutsamer Unterschied zu den beiden Zeitpunkten. Zwar sind auch hier die Mittelwerte hypothesenkonform immer geringer, wenn die Versuchspersonen das erste Mal ihre erfundene Unfallserie wiedergeben sollen, als wenn sie bereits eine Befragung hinter sich hatten, allerdings sind diese Veränderungen wie gesagt nicht signifikant.

Des Weiteren haben hier offenbar wieder die Fehlerzahlen Gewicht. Sowohl bezüglich der Gesamtfehlerzahl als auch der Zahl der allgemeinen Erinnerungsfehler scheinen die Versuchspersonen in ihren selbst erfundenen Unfallserien bei einer ersten Befragung nach acht Wochen deutlich mehr Probleme zu haben. Die beiden Fehlerraten erweisen sich zu NT2 als deutlich höher als zu T2 und im statistischen Ergebnis als hochsignifikant verschieden ($T(20) = 3,077$; $p = 0,003$ bzw. $T(20) = 2,738$; $p = 0,007$).

Den Quellenverwechslungen kommt auch hier wiederum keine Bedeutung zu. Es zeigen sich keinerlei signifikante Veränderungen zwischen den Zeitpunkten nach ca. acht Wochen.

Ein statistisch ebenso unbedeutender Unterschied ist hinsichtlich der genannten Konfabulationen zu beobachten. Allerdings ist der Mittelwert zu NT2 etwas höher als zu T2.

Ein innerhalb einer kleinen Ober- und Untergrenze aufzufindender, aber doch signifikanter Unterschied zwischen den Befragungszeitpunkten besteht für das Vorkommen schematypischer Intrusionen. Diese sind im Mittel bei einer ersten Befragung nach acht Wochen um 0,75 Punkte höher als bei einer zweiten ($T(20) = 1,872$; $p = 0,038$).

Die letzten der bekannten Übersichtstabellen zu den genauen Werten und Ergebnissen finden sich im Anschluss (Tabellen 4-60 und 4-61).

Tabelle 4-60: Mittelwerte und Standardabweichungen aller abhängigen Variablen zu T2 und NT2 für die 'reality monitoring' – Gruppe. Die Angaben beziehen sich auf Serie 2 - analog Dia-Serie 'PKW-Fahrrad'.

Variable	T2		NT2	
	MW	SD	MW	SD
korrekte Details gesamt	12,83	9,32	9,80	5,25
korrekte Details im 'free recall'	5,75	3,79	3,50	3,21
korrekte Details 'auf Nachfrage'	7,08	7,93	6,30	3,80
Fehler gesamt	29,17	6,65	39,10	8,50
Quellenverwechslungen insgesamt	1,17	0,94	1,40	1,65
Quellenverwechslungen innerhalb	0,58	0,90	0,70	0,82
Quellenverwechslungen zwischen	0,58	0,79	0,70	1,06
Schematypische Intrusionen	0,25	0,62	0,90	0,99
Konfabulationen	4,08	2,07	3,10	2,33
allgemeine Erinnerungsfehler	23,67	7,83	33,70	9,37

Tabelle 4-61: Ergebnisse der t-Tests aller abhängigen Variablen, getrennt für die beiden Unfall - Serien für die 'reality monitoring' – Gruppe zu T2 und NT2

Variable	SERIE 1 (PKW-PKW)			SERIE 2 (PKW-Fahrrad)		
	T	df	p	T	df	p
korrekte Details gesamt	-0,434	20	0,335	-0,913	20	0,186
korrekte Details im 'free recall'	-0,682	20	0,252	-1,484	20	0,077
korrekte Details 'auf Nachfrage'	-0,063	20	0,475	-0,286	20	0,389
Fehler gesamt	2,000	20	0,030	3,077	20	0,003
Quellenverwechslungen gesamt	1,310	15,283	0,105	0,418	20	0,341
Quellenverwechslungen innerhalb	1,310	15,283	0,105	0,314	20	0,378
Quellenverwechslungen zwischen	<i>n. berechenbar, MWe = 0</i>			0,295	20	0,386
Schematypische Intrusionen	3,320	20	0,002	1,872	20	0,038
Konfabulationen	0,469	20	0,322	-1,049	20	0,154
allgemeine Erinnerungsfehler	1,270	20	0,110	2,738	20	0,007

Tabelle 4-62: zusammenfassende Darstellung der Ergebnisse aller *t* – Tests der Veränderung der Gedächtnisleistungen über die Zeit und zu den Befragungszeitpunkten für alle Variablen für beide Gruppen sowie beide Dia- bzw. Unfall - Serien zusammen ('PKW-PKW', 'PKW-Fahrrad') und getrennt voneinander ('PKW-PKW' bzw. 'PKW-Fahrrad'); wobei: P = PKW und F = Fahrrad.

Variable	Gruppe: SOURCE MONITORING									Gruppe: REALITY MONITORING								
	T1-T2			T1-NT2			T2-NT2			T1-T2			T1-NT2			T2-NT2		
	P-P P-F	P-P	P-F	P-P P-F	P-P	P-F	P-P P-F	P-P	P-F	P-P P-F	P-P	P-F	P-P P-F	P-P	P-F	P-P P-F	P-P	P-F
korrekte Details gesamt				*		***	**		**	*		**			**			
korrekte Details im 'free recall'				**		**	***	**	**						*			
korrekte Details 'auf Nachfrage'						*				*		*			*			
Fehler gesamt	***		**							***	**	***	***	***	***	***	*	**
Quellenverwechslungen gesamt																		
Quellenverwechslungen innerhalb der Dia-Serien																		
Quellenverwechslungen zwischen den Serien																		
schematypische Intrusionen				**		**	**		**		*		*	*		**	**	*
Konfabulationen				*						*		*						
allgemeine Erinnerungsfehler	**		*							***	**	***	***	***	***	***		**

wobei: ' * ' = signifikantes Ergebnis ($p \leq 0,05$)

' ** ' = hochsignifikantes Ergebnis ($p \leq 0,01$)

' *** ' = höchstsignifikantes Ergebnis ($p \leq 0,001$)

' ' = kein signifikantes Ergebnis ($p > 0,05$)

4.1.5 Ergebnisse zu den Interaktionseffekten

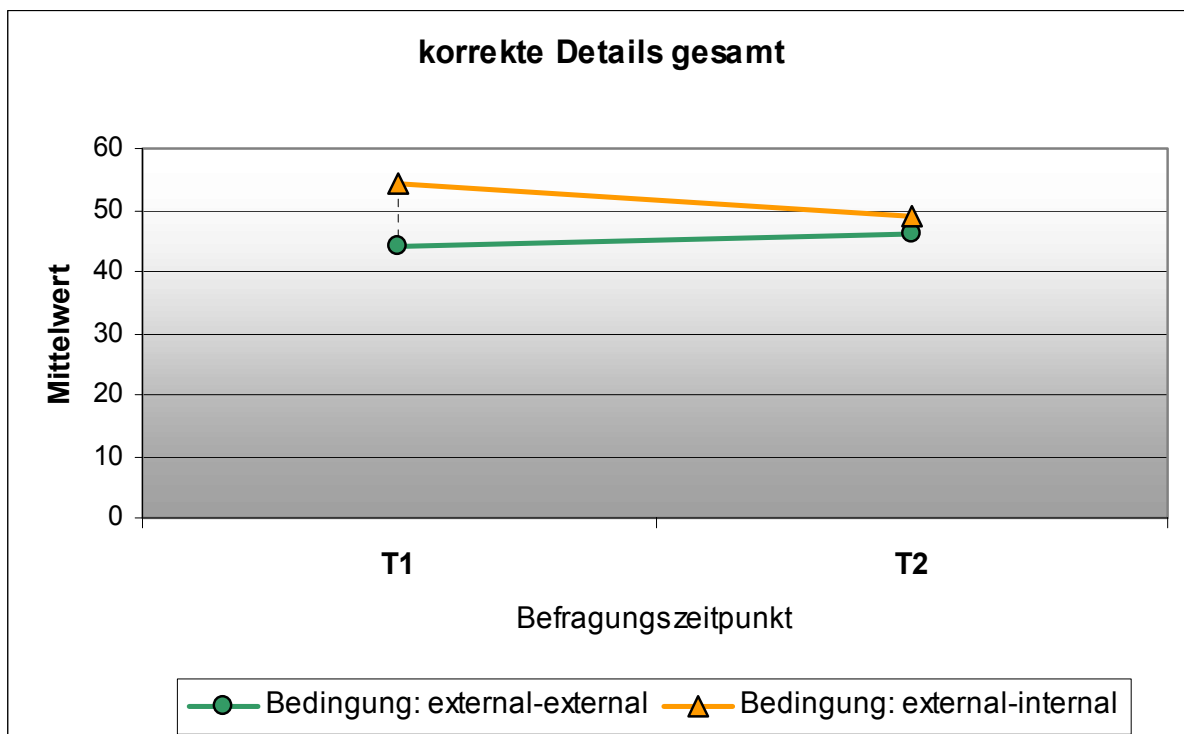
Anhand des 'Allgemeinen Linearen Modells (ALM) mit Messwiederholung' wurde überprüft, ob sich für die in den Aussagen berichteten korrekten Details ('gesamt', 'free recall' und 'auf Nachfrage') sowie für die verschiedenen Fehlerarten signifikante Veränderungen in der Ausprägung in Abhängigkeit von der Gruppenzugehörigkeit (Versuchsbedingung) vom ersten zum zweiten Befragungszeitpunkt in Form einer statistischen Wechselwirkung (Interaktion) zeigen. Zudem sind die jeweiligen gegebenenfalls vorhandenen Haupteffekte 'Gruppe' und 'Befragungszeitpunkt' dargestellt.

Dazu wurden in die Analyse alle diejenigen Versuchspersonen miteinbezogen, die zu beiden Befragungszeitpunkten (T1 und T2) interviewt wurden ($N = 23$). Die externe 'source monitoring' – Gruppe umfasste dabei $n = 11$ Versuchspersonen, die 'reality monitoring' – Gruppe $n = 12$.

Im Einzelnen ergaben sich dabei folgende Befunde:

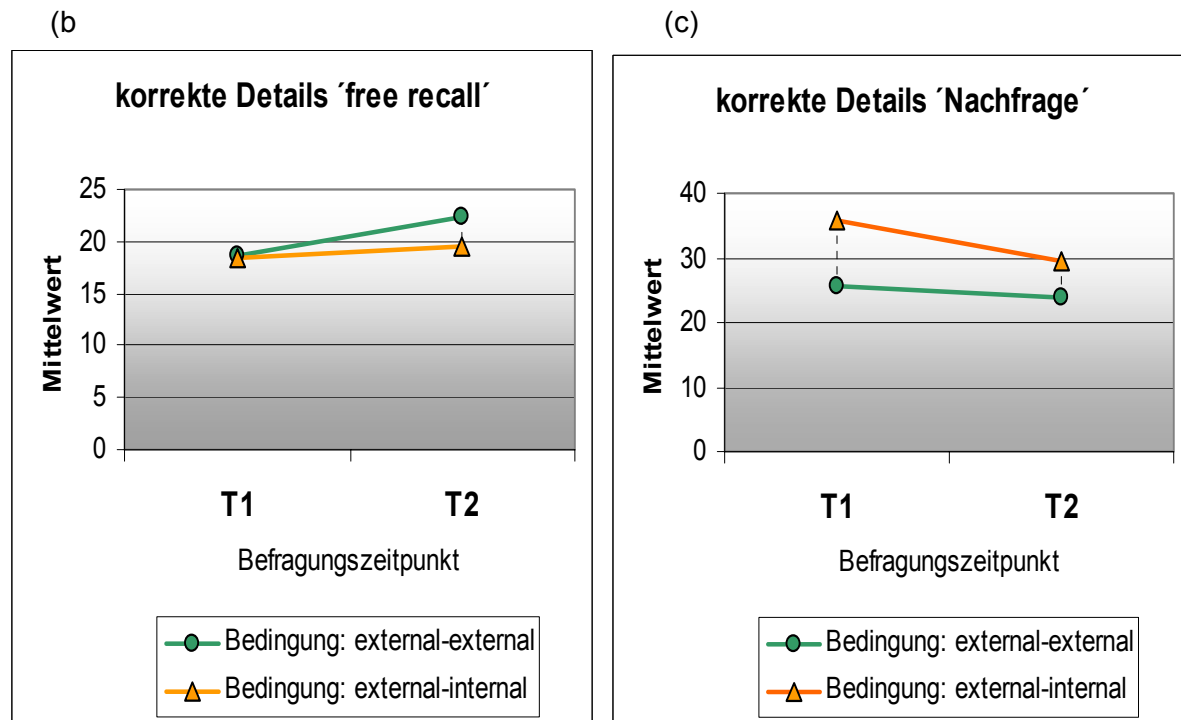
Abbildungen 4-22-a.b.c: Graphische Darstellung der Haupt- und Interaktionseffekte bezüglich der Variablen 'korrekte Details gesamt' (a), 'korrekte Details im free recall' (b) und 'korrekte Details auf Nachfrage' (c)

(a)



Für die insgesamt korrekt berichteten Details ergibt sich hinsichtlich des Faktors 'Versuchsbedingung' kein signifikanter Haupteffekt ($F(1;21) = 1,381$; $p = 0,253$). Hinsichtlich

des Befragungszeitpunktes findet sich ebenfalls kein signifikanter Haupteffekt ($F(1;21) = 0,446$; $p = 0,512$). Eine signifikante Wechselwirkung zwischen den beiden Faktoren kann ebenso nicht festgestellt werden ($F(1;21) = 3,054$; $p = 0,095$). Entsprechend klein sind die Werte der jeweiligen Effektstärken (siehe ggf. Tabelle 4-66).



Ähnliche Befunde ergeben sich für die korrekt berichteten Details im 'free recall' und auch auf Nachfrage:

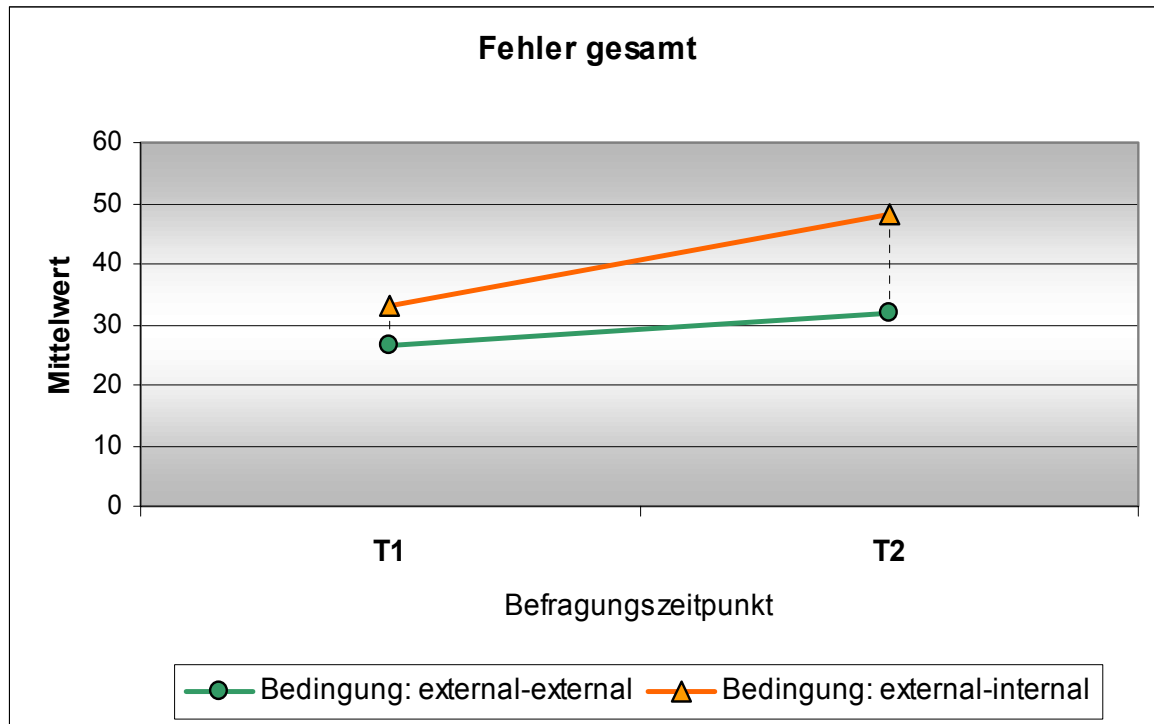
Für erstgenannte findet sich mit $F(1;21) = 0,201$ und $p = 0,658$ weder ein signifikanter Haupteffekt 'Gruppe' noch ein signifikanter Haupteffekt 'Befragungszeitpunkt' ($F(1;21) = 2,419$; $p = 0,135$). Auch eine statistisch bedeutsame Interaktion zwischen den Faktoren kann nicht festgestellt werden ($F(1;21) = 0,731$; $p = 0,402$).

Ebenso verhält es sich im Ergebnis für die auf Nachfrage korrekt berichteten Details. Auch hier lässt sich – wenn auch gerade nicht mehr – kein signifikanter Haupteffekt 'Gruppe' gefunden werden. Das Signifikantniveau wird hier nur denkbar knapp verfehlt ($F(1;21) = 4,266$; $p = 0,051$). Ein signifikanter Haupteffekt 'Befragungszeitpunkt' ($F(1;21) = 3,320$; $p = 0,083$) sowie eine signifikante Interaktion ($F(1;21) = 1,269$; $p = 0,273$) ergeben sich ebenfalls nicht.

Somit ergeben sich für alle Variablen bezüglich der Haupteffekte und Interaktion sehr schwache Werte für die Effektstärken (siehe Tabelle 4-66).

Abbildungen 4-23-a.b.c.d.e.f.g: Graphische Darstellung der Haupt- und Interaktionseffekte bezüglich der Fehler gesamt (a) und der Fehlerarten [Quellenverwechslungen (b,c,d), schematypische Intrusionen (e), Konfabulationen (f) und allgemeine Erinnerungsfehler (g)]

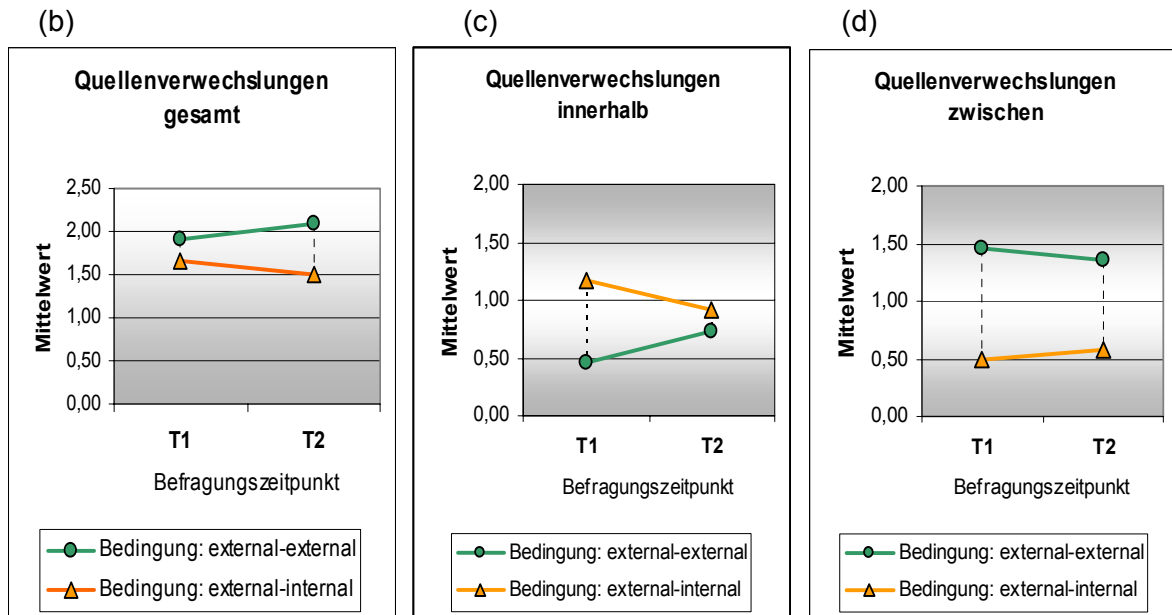
(a)



Bei Betrachtung der Gesamtfehlerzahl zeigen sich jeweils höchstsignifikante Ergebnisse sowohl für die beiden Haupteffekte als auch für die Interaktion zwischen den beiden Faktoren.

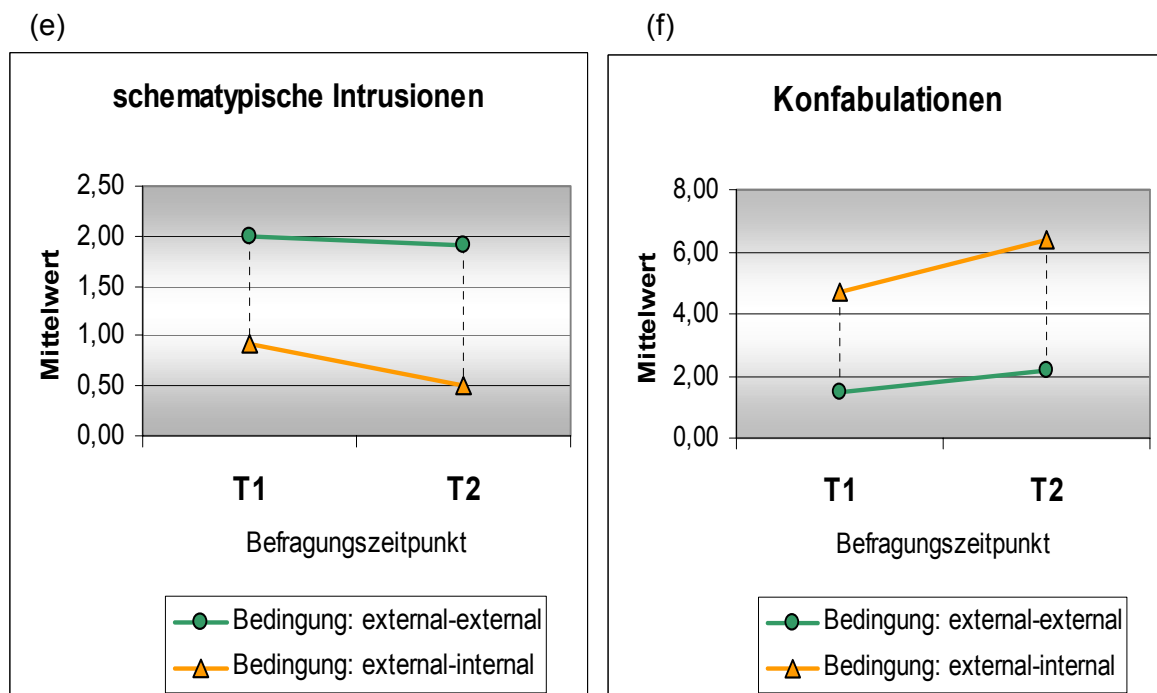
Für den Faktor 'Versuchsbedingung' ergibt sich mit $F(1;21) = 24,987$ und $p < 0,001$ ein höchstsignifikanter F-Wert, ebenso für den Faktor 'Befragungszeitpunkt' ($F(1;21) = 120,892$; $p < 0,001$). Mit einer mittleren bzw. hohen Effektstärke von $d = 0,543$ bzw. $d = 0,852$ kann hier auch von einer praktischen und praxisrelevanten Bedeutsamkeit der jeweiligen Unterschiede ausgegangen werden.

Auch für die Interaktion zwischen den beiden Faktoren zeigt die ALM-Analyse ein höchstsignifikantes Ergebnis ($F(1;21) = 28,080$; $p < 0,001$) und mit $d = 0,572$ auch einen mittelstarken Effekt.



Für die Formen und Variablen der Quellenverwechslung gilt:

Die Veränderungen in den Versuchsbedingungen sowie zu den Befragungszeitpunkten unterscheiden sich nicht wesentlich. Es konnten mit Ausnahme eines Haupteffekts 'Gruppe' für die Quellenverwechslungen zwischen den Serien ($F(1;21) = 5,619$; $p = 0,027$) jeweils weder signifikante Haupteffekte noch Interaktionen beobachtet werden. Der genannte Haupteffekt weist eine – jedoch geringe – Effektgröße von $d = 0,211$ auf.

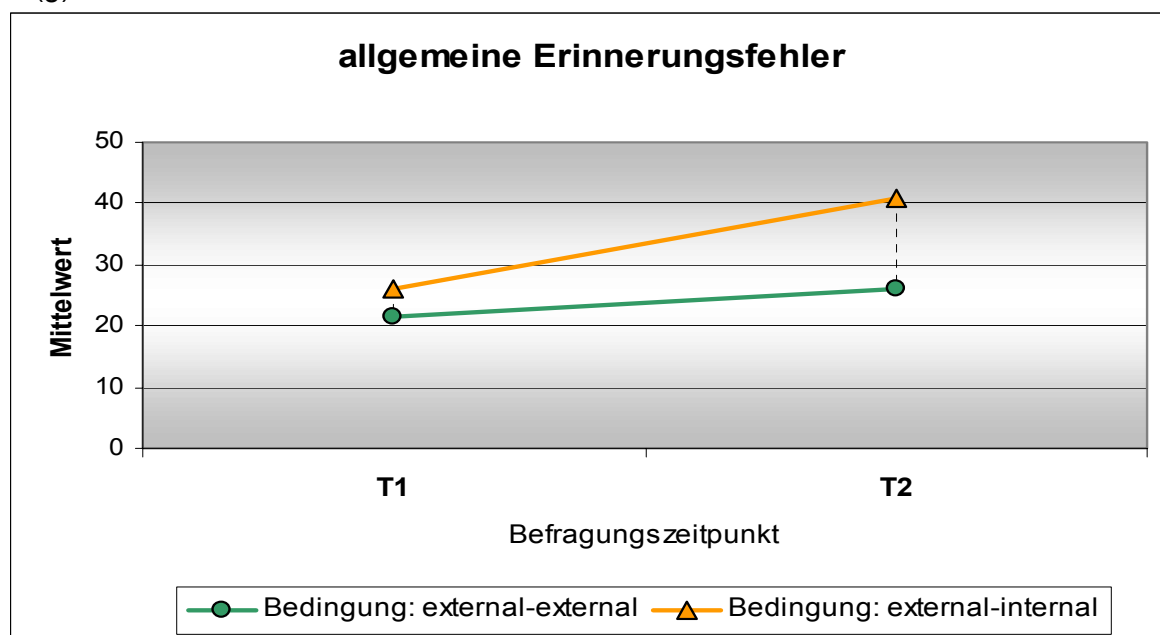


Hinsichtlich der schematypischen Intrusionen ergibt sich für die Gruppenzugehörigkeit ein hochsignifikanter Wert ($F(1;21) = 12,948$; $p = 0,002$), hingegen weist das ALM sowohl für

den Befragungszeitpunkt ($F(1;21) = 0,867$; $p = 0,362$) als auch für die Wechselwirkung zwischen den beiden Faktoren ($F(1;21) = 0,357$; $p = 0,557$) keine signifikanten Werte aus. Mit $d = 0,381$ bezüglich des Haupteffekts 'Gruppe' kann hier von einer geringen bis mittleren Effektstärke gesprochen werden.

Was die Variable 'Konfabulationen' betrifft, zeigt sich für den Faktor 'Versuchsbedingung' mit $F(1;21) = 16,987$ und $p < 0,001$ ein höchstsignifikantes Ergebnis. Diesbezüglich ebenfalls als statistisch bedeutsam erweist sich mit einem signifikanten Ergebnis der Faktor 'Befragungszeitpunkt' ($F(1;21) = 5,502$; $p = 0,029$). In der weiteren Analyse kann jedoch keine signifikante Veränderung zwischen den Versuchsgruppen in Abhängigkeit der beiden Befragungszeitpunkte festgestellt werden – eine Interaktion liegt also nicht vor ($F(1;21) = 0,938$; $p = 0,344$). Der Haupteffekt 'Versuchsbedingung' weist mit $d = 0,447$ einen doch mittleren Effekt auf, der Haupteffekt 'Befragungszeitpunkt' mit $d = 0,208$ einen eher geringen.

(g)



Die allgemeinen Erinnerungsfehler erweisen sich als sehr gutes Kriterium zur Differenzierung sowohl zwischen den Versuchsbedingungen als auch zwischen den Erhebungszeitpunkten. Für beide Faktoren ergeben sich höchstsignifikante Unterschiede ($F(1;21) = 18,804$; $p < 0,001$ bzw. $F(1;21) = 78,931$; $p < 0,001$) und entsprechend auch mittlere bzw. hohe Effektstärken ($d = 0,472$ bzw. $d = 0,790$). Für die Wechselwirkung zwischen Gruppenzugehörigkeit und Befragungszeitpunkt weist das ALM mit $F(1;21) = 29,983$ und $p < 0,001$ ebenfalls einen höchstsignifikanten Wert aus und ergänzt diesen Befund durch einen (mittleren) Effektgrößenwert von $d = 0,533$. Diese Ergebnisse haben also mit den gegebenen Effektstärken somit auch durchaus praxisrelevante Bedeutung.

Den *Tabellen 4-63 bis 4-65* sind die entsprechenden genauen F-, df- und Signifikanzwerte des ALM bezüglich der Haupteffekte sowie der Interaktion für alle abhängigen Variablen zu entnehmen.

Tabelle 4-66 stellt eine Übersicht der Effektstärkenindizes für die Haupteffekte sowie für die Interaktion der Faktoren, ebenfalls für alle abhängigen Variablen, dar.

Tabelle 4-63 : Ergebnisse des ALM mit Messwiederholung bezüglich 'Haupteffekt Gruppe' für alle untersuchten Variablen ('source- / reality monitoring')

Variable	F	df	Signifikanz
korrekte Details gesamt	1,381	1;21	0,253
korrekte Details im 'free recall'	0,201	1;21	0,658
korrekte Details 'auf Nachfrage'	4,266	1;21	0,051
Fehler gesamt	24,987	1;21	< 0,001
Quellenverwechslungen insgesamt	0,763	1;21	0,392
Quellenverwechslungen innerhalb	1,429	1;21	0,245
Quellenverwechslungen zwischen	5,619	1;21	0,027
schematypische Intrusionen	12,948	1;21	0,002
Konfabulationen	16,987	1;21	< 0,001
allgemeine Erinnerungsfehler	18,804	1;21	< 0,001

Tabelle 4-64 : Ergebnisse des ALM mit Messwiederholung bezüglich 'Haupteffekt Befragungszeitpunkt' für alle untersuchten Variablen ('source- / reality monitoring')

Variable	F	df	Signifikanz
korrekte Details gesamt	0,446	1;21	0,512
korrekte Details im 'free recall'	2,419	1;21	0,135
korrekte Details 'auf Nachfrage'	3,320	1;21	0,083
Fehler gesamt	120,892	1;21	< 0,001
Quellenverwechslungen insgesamt	0,000	1;21	0,986
Quellenverwechslungen innerhalb	0,001	1;21	0,970
Quellenverwechslungen zwischen	0,000	1;21	0,988
schematypische Intrusionen	0,867	1;21	0,362
Konfabulationen	5,502	1;21	0,029
allgemeine Erinnerungsfehler	78,931	1;21	< 0,001

Tabelle 4-65 : Ergebnisse des ALM mit Messwiederholung bezüglich 'Interaktion Gruppe – Befragungszeitpunkt' für alle untersuchten Variablen ('source- / 'reality monitoring')

Variable	F	df	Signifikanz
korrekte Details gesamt	3,054	1;21	0,095
korrekte Details im 'free recall'	0,731	1;21	0,402
korrekte Details 'auf Nachfrage'	1,269	1;21	0,273
Fehler gesamt	28,080	1;21	< 0,001
Quellenverwechslungen insgesamt	0,172	1;21	0,683
Quellenverwechslungen innerhalb	0,776	1;21	0,388
Quellenverwechslungen zwischen	0,123	1;21	0,730
schematypische Intrusionen	0,357	1;21	0,557
Konfabulationen	0,938	1;21	0,344
allgemeine Erinnerungsfehler	23,983	1;21	< 0,001

Tabelle 4-66: Effektstärken (1) - 'source- / 'reality monitoring'

Variable	Effektstärkenindex d		
	Haupteffekt 'GRUPPE'	Haupteffekt 'BEFRAGUNGS- ZEITPUNKT'	Interaktion 'GRUPPE – BEFRAGUNGS- ZEITPUNKT'
korrekte Details gesamt	0,062	0,021	0,127
korrekte Details im 'free recall'	0,009	0,103	0,034
korrekte Details 'auf Nachfrage'	0,169	0,137	0,057
Fehler gesamt	0,543	0,852	0,572
Quellenverwechslungen insgesamt	0,035	0,000	0,008
Quellenverwechslungen innerhalb	0,064	0,000	0,036
Quellenverwechslungen zwischen	0,211	0,000	0,006
schematypische Intrusionen	0,381	0,040	0,017
Konfabulationen	0,447	0,208	0,043
allgemeine Erinnerungsfehler	0,472	0,790	0,533

Die nachstehende Tabelle 4-67 fasst nochmals alle Ergebnisse des ALM mit Messwiederholung zu den Haupteffekten und Interaktionen zusammen.

Tabelle 4-67: zusammenfassende Darstellung aller Ergebnisse des ALM mit Messwiederholung für alle Variablen der 'source-' / 'reality monitoring' - Fragestellung

Variable	Haupteffekt 'GRUPPE'	Haupteffekt 'BEFRAGUNGS- ZEITPUNKT'	Interaktion 'GRUPPE – BEFRAGUNGS- ZEITPUNKT'
korrekte Details gesamt			
korrekte Details im 'free recall'			
korrekte Details 'auf Nachfrage'			
Fehler gesamt	***	***	***
Quellenverwechslungen insgesamt			
Quellenverwechslungen innerhalb			
Quellenverwechslungen zwischen	*		
schematypische Intrusionen	**		
Konfabulationen	***	*	
allgemeine Erinnerungsfehler	***	***	***

wobei:

' ' = kein signifikantes Ergebnis ($p > 0,05$) ' * ' = signifikantes Ergebnis ($p \leq 0,05$)

' ** ' = hochsignifikantes Ergebnis ($p \leq 0,01$) ' *** ' = höchstsignifikantes Ergebnis ($p \leq 0,001$)

4.2 Ergebnisse zur Glaubhaftigkeit v. Zeugenaussagen (Validierungsuntersuchung)

Im Folgenden werden die Ergebnisse zur Validierungsstudie der 17 (19) in die statistische Analyse einbezogenen Aussagemerkmale (Realkennzeichen) im einzelnen sowie bezüglich der fünf Merkmalskategorien als auch aller Aussagemerkmale insgesamt dargestellt.

Nachdem zunächst die mittels des Einstichproben-KOLMOGOROV-SMIRNOV-Test für jedes Realkennzeichen einzeln sowie die Merkmalskomplexe überprüfte Verteilungsform sowohl bezüglich des Wahrheitsstatus als auch des Befragungszeitpunktes tabellarisch dargestellt wurde, folgen im Anschluss die Ausführungen für o.g. Ergebnisse.

Dabei werden jeweils graphisch die Gruppenmittelwerte und Standardabweichungen sowie die Darstellungen der gegebenenfalls gefundenen Haupt- und Interaktionseffekte angeführt. Dies wurde aus Gründen der Vollständigkeit und der Übersichtlichkeit auch für den Fall nicht vorhandener Effekte oder Wechselwirkungen beschrieben.

Im Anschluss folgen die statistische Ergebnisdarstellung der für alle Aussagemerkmale einzeln sowie deren Merkmalskategorien und alle Aussagemerkmale insgesamt berechneten univariaten Varianzanalyse mit Messwiederholung (Allgemeines Lineares Modell, im folgenden 'ALM').

Zur weiteren Differenzierung wurden jeweils post-hoc als parametrisches Verfahren der t-Test für abhängige Stichproben bzw. als nicht-parametrisches Verfahren der Wilcoxon-Test für gepaarte Differenzen gerechnet, je nachdem, ob beim zu analysierenden Vergleich des Wahrheitsstatus bzw. des Befragungszeitpunktes unter- bzw. miteinander die Daten einer Normalverteilung entsprachen oder nicht. War eine der zwei zu vergleichenden Ausprägung der Variablen (wT1, eT1, wT2, eT2) nicht normalverteilt, wurde post-hoc das genannte verteilungsfreie Verfahren gerechnet.

Die angegebenen p-Werte des K-S-Tests sowie des ALM beziehen sich auf eine zweiseitige Hypothesenprüfung, die Signifikanzergebnisse der post-hoc-Verfahren beruhen auf einer einseitigen Hypothesenprüfung.

Sowohl hinsichtlich der Überprüfung der Verteilungsform als auch aller angewendeten Signifikanzverfahren wurde von einer Irrtumswahrscheinlichkeit von 5 % ($p = 0,05$) ausgegangen.

Die diesbezüglichen Ergebnisse werden dann noch tabellarisch zusammengefasst.

Abschließend folgt die Ergebnisdarstellung zur Einschätzung der Glaubhaftigkeit der Aussagen sowie zur Klassifikation der Glaubhaftigkeit mittels Diskriminanzanalyse.

4.2.1 Überprüfung auf Verteilungsform

Tabelle 4-68: Ergebnisse des K-S-Tests auf Normalverteilung (3) für alle Aussagemerkmale und Merkmalskategorien

Variable bzw. Merkmalskategorie	asymptotische Signifikanz (zweiseitig)			
	wT1	eT1	wT2	eT2
Logische Konsistenz	0,006	0,032	0,037	0,036
Widerspruchsfreie, ungeordnete Darstellung	0,039	0,012	0,069	0,059
Quantitativer Detailreichtum	0,096	0,132	0,081	0,096
Kontextuelle Einbettung	0,105	0,015	0,265	0,075
Schilderung nonverbaler Aktionen	0,113	0,004	0,019	0,017
Wiedergabe von Gesprächen	0,304	0,279	0,178	0,233
Schilderung von Komplikationen i. Handlungsverl.	0,074	0,005	0,045	< 0,001
Ausgefallene Details	0,001	0,001	0,001	< 0,001
Nebensächliche Details	0,029	0,093	0,003	0,045
Schilderung unverstandener Handlungselemente	-	-	< 0,001	-
Inhaltliche Verschachtelungen	< 0,001	< 0,001	< 0,001	< 0,001
Schilderung eigener psych. od. phys. Vorgänge	0,198	0,003	0,377	0,004
Sch. psych. od. phys. V. beim Täter / Beteiligten	0,043	0,002	0,002	0,001
Spontane Verbesserungen der eigenen Aussage	0,233	0,023	0,018	0,016
Zugeben von Erinnerungslücken	0,007	0,081	0,098	0,024
Einwände gegen die Richtigkeit der eig. Aussage	0,001	< 0,001	< 0,001	< 0,001
Selbstbelastungen	0,082	0,010	0,002	0,002
Inschutznahme des Täters oder Beteiligten	< 0,001	< 0,001	< 0,001	< 0,001
Deliktspezifische Aussageelemente	< 0,001	< 0,001	< 0,001	-
Allgemeine Merkmale	0,554	0,473	0,498	0,509
Spezielle Inhalte	0,970	0,271	0,420	0,201
Inhaltliche Besonderheiten	0,102	0,462	0,230	0,178
Motivationsbezogene Inhalte	0,286	0,135	0,082	0,352
Deliktspezifische Inhalte	< 0,001	< 0,001	< 0,001	-
RKZ total	0,663	0,821	0,720	0,787

wobei: wT1/ wT2: Ausprägung der Variable der erlebnis-fundierten Geschichte zu T1 bzw. T2

eT1/ eT2: Ausprägung der Variable der nicht erlebnis-fundierten Geschichte zu T1/T2

Dabei kann bei den **fett markierten** Ergebnissen der jeweiligen Variablenausprägungen von einer Normalverteilung der Daten ausgegangen werden, bei den anderen nicht.

Aus Gründen der Übersichtlichkeit und Vollständigkeit wurde ebenfalls für jedes Aussage-merkmal (Realkennzeichen) bzw. jede Merkmalskategorie eine Analyse mittels des ALM durchgeführt, auch wenn die Annahmen einer Normalverteilung nicht unbedingt gegeben waren.

Jedoch sei an dieser Stelle kurz auf das zentrale Grenzwerttheorem verwiesen, wonach die Verteilung von Mittelwerten aus Stichproben des Umfangs 'n', die sämtlich derselben Grundgesamtheit entnommen wurden, mit wachsendem Stichprobenumfang in eine Normalverteilung übergeht (approximative Normalverteilung). Da für praktische Zwecke davon ausgegangen wird, dass die Mittelwertverteilung für beliebige Verteilungsformen eines Merkmals in der Population bei $n \geq 30$ bereits hinreichend normalverteilt ist (vgl. BORTZ, 1999), soll dies auch in der vorliegenden Arbeit (mit $n = 25$) der Fall sein, was die Berechnungen des ALM betrifft.

Jedoch wurden zur genauen Analyse post-hoc - wie in 4.2 bereits erläutert – gegebenenfalls parametrische bzw. verteilungsfreie Verfahren gerechnet.

Aufgrund der sehr geringen bzw. fast und gar nicht vorhandenen Mittelwerte und entsprechender Verteilung der Daten wurden die Realkennzeichen '(Phänomengemäße) Schilderung unverstandener Handlungselemente' und 'Inhaltliche Verschachtelungen' nicht in die weitere statistische Analyse miteinbezogen.

4.2.2 Ergebnisse zu den einzelnen Aussagemerkmalen

4.2.2.1 Logische Konsistenz

Abbildung 4-24-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Logische Konsistenz

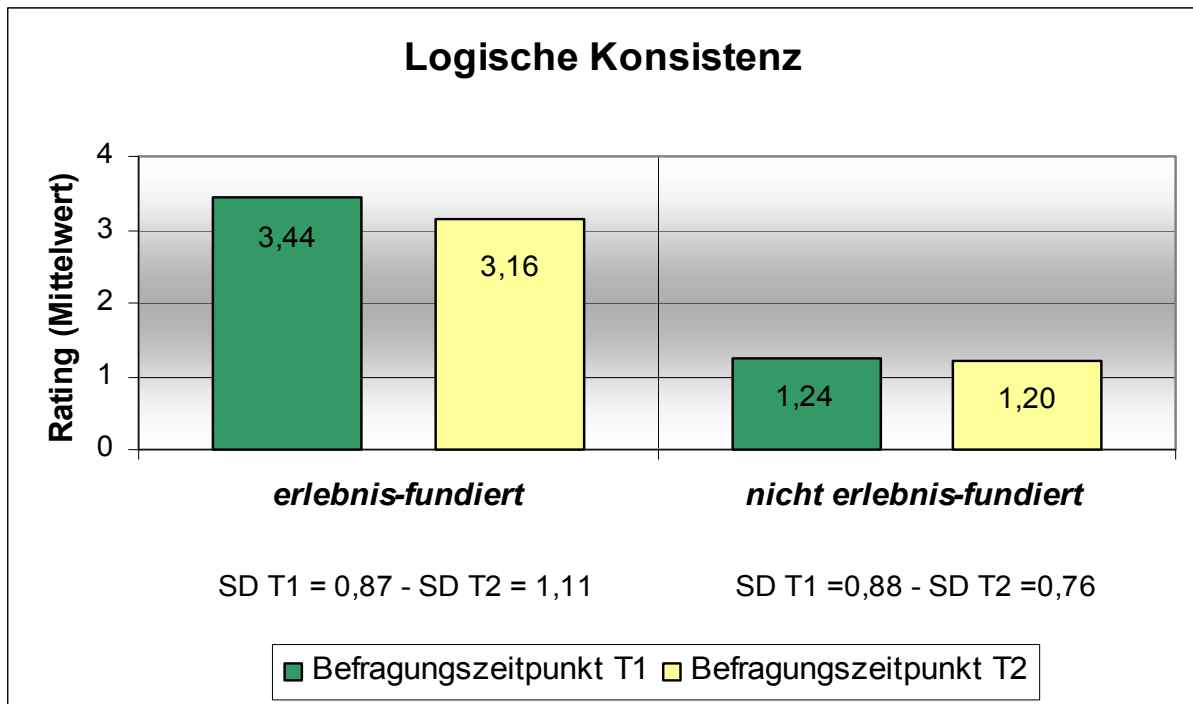
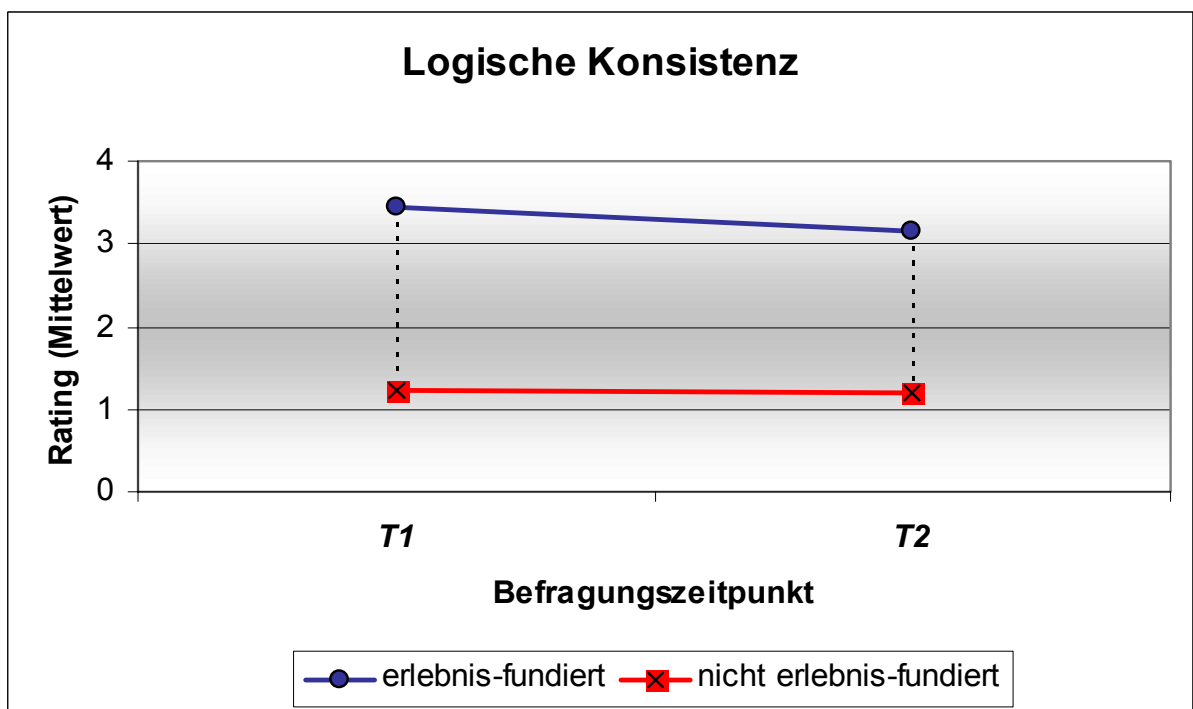


Abbildung 4-24-2: Haupt- und Interaktionseffekte zu T1 und T2: Logische Konsistenz



Die univariate ALM-Analyse ergibt für das wichtige Aussagemerkmal 'Logische Konsistenz' einen höchstsignifikanten Haupteffekt 'Wahrheitsstatus' ($F(1;24) = 56,017$ und $p < 0,001$). Jedoch konnte bezüglich des Faktors 'Befragungszeitpunkt' kein signifikanter Unterschied nachgewiesen werden ($F(1;24) = 2,621$ und $p = 0,119$).

Ebenso fand sich mit $F(1;24) = 0,658$ und $p = 0,425$ keine signifikante Interaktion zwischen den beiden Faktoren.

Anhand des post-hoc durchgeführten Wilcoxon-Test konnte weiterhin zu beiden Befragungszeitpunkten T1 und T2 ein höchstsignifikanter Unterschied in der Merkmalsausprägung zwischen wahren und erfundenen Geschichten ermittelt werden ($z = -3,900$ und $p < 0,001$ bzw. $z = -3,827$ und $p < 0,001$). Die entsprechenden Mittelwerte waren für die erfundenen Geschichten wesentlich geringer als für die wahren Geschichten.

Ebenfalls mittels des Wilcoxon-Test ließen sich post-hoc allerdings weder für die wahren noch für die erfundenen Geschichten statistisch bedeutsame Veränderungen vom ersten zum zweiten Befragungszeitpunkt feststellen ($z = -1,403$ und $p = 0,081$ bzw. $z = -0,258$ und $p = 0,398$).

4.2.2.2 Widerspruchsfreie, ungeordnete Darstellung

Abbildung 4-25-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Widerspruchsfreie, ungeordnete Darstellung

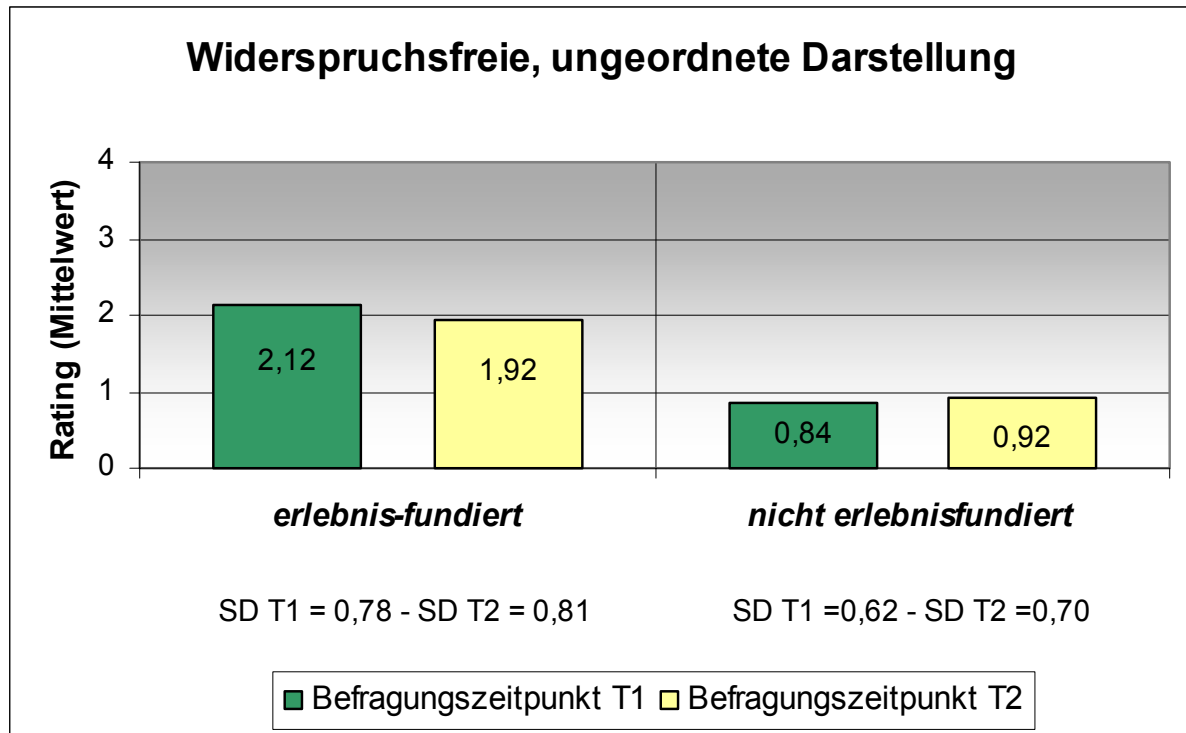
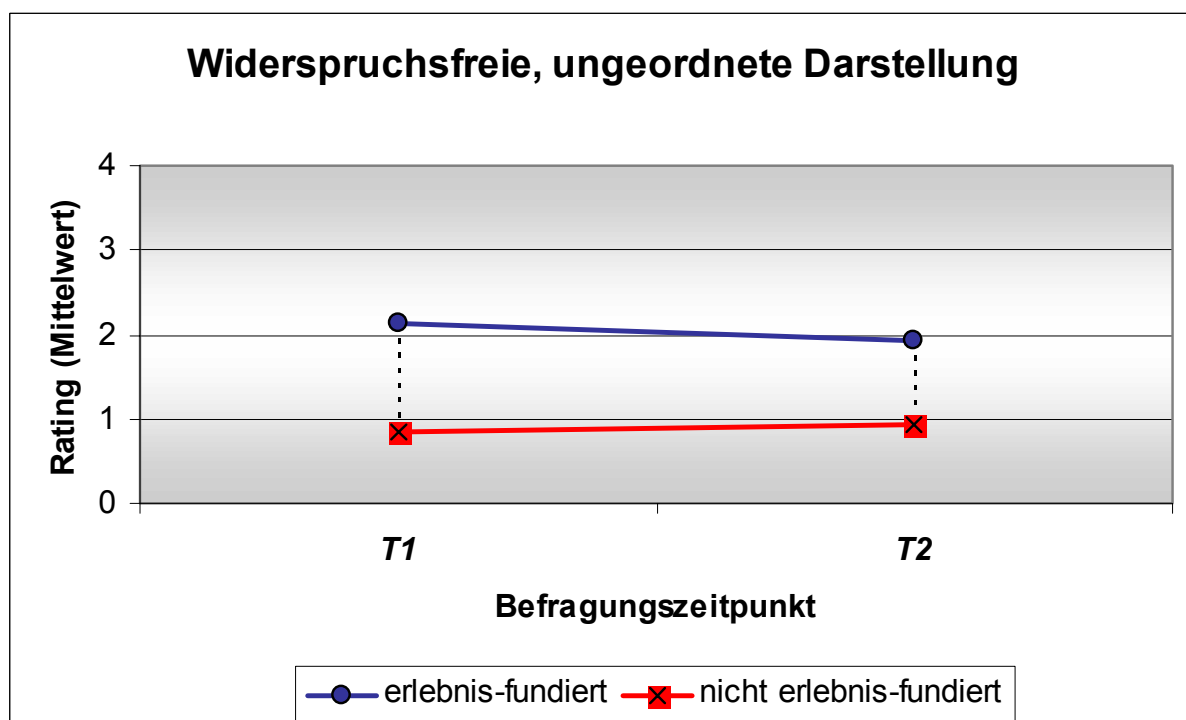


Abbildung 4-25-2: Haupt- und Interaktionseffekte zu T1 und T2: Widerspruchsfreie, ungeordnete Darstellung



Wie für das voranstehende Realkennzeichen fand sich auch für dieses ein deutlicher Unterschied bezüglich des Faktors 'Wahrheitsstatus' in Form eines ebenfalls höchstsignifikanten Haupteffekts ($F(1;24) = 37,561$ und $p < 0,001$).

Ebenso analog gestalten sich die Befunde hinsichtlich des Befragungszeitpunktes sowie einer eventuellen Wechselwirkung zwischen 'Wahrheitsstatus' und 'Befragungszeitpunkt'. Auch hier konnten jeweils keine signifikanten Unterschiede nachgewiesen werden ($F(1;24) = 0,302$ und $p = 0,588$ bzw. $F(1;24) = 1,205$ und $p = 0,283$).

Bei genauerer post-hoc-Untersuchung des Haupteffekts 'Status' ergab der Wilcoxon-Test zum ersten Befragungszeitpunkt T1 einen höchstsignifikanten Unterschied ($z = -3,909$ und $p < 0,001$) zwischen den Ratings der beiden Geschichten, ebenso wie der t-Test für abhängige Stichproben zum zweiten Befragungszeitpunkt T2 ($t(24) = 3,974$ und $p < 0,001$).

Hinsichtlich des mittels Wilcoxon-Test post-hoc detaillierter untersuchten Befragungszeitpunktes zeigten sich allerdings weder für die wahren noch für die erfundenen Aussagen signifikante Unterschiede zwischen T1 und T2 ($z = -1,000$ und $p = 0,159$ bzw. $z = -0,587$ und $p = 0,279$).

4.2.2.3 Quantitativer Detailreichtum

Abbildung 4-26-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Quantitativer Detailreichtum

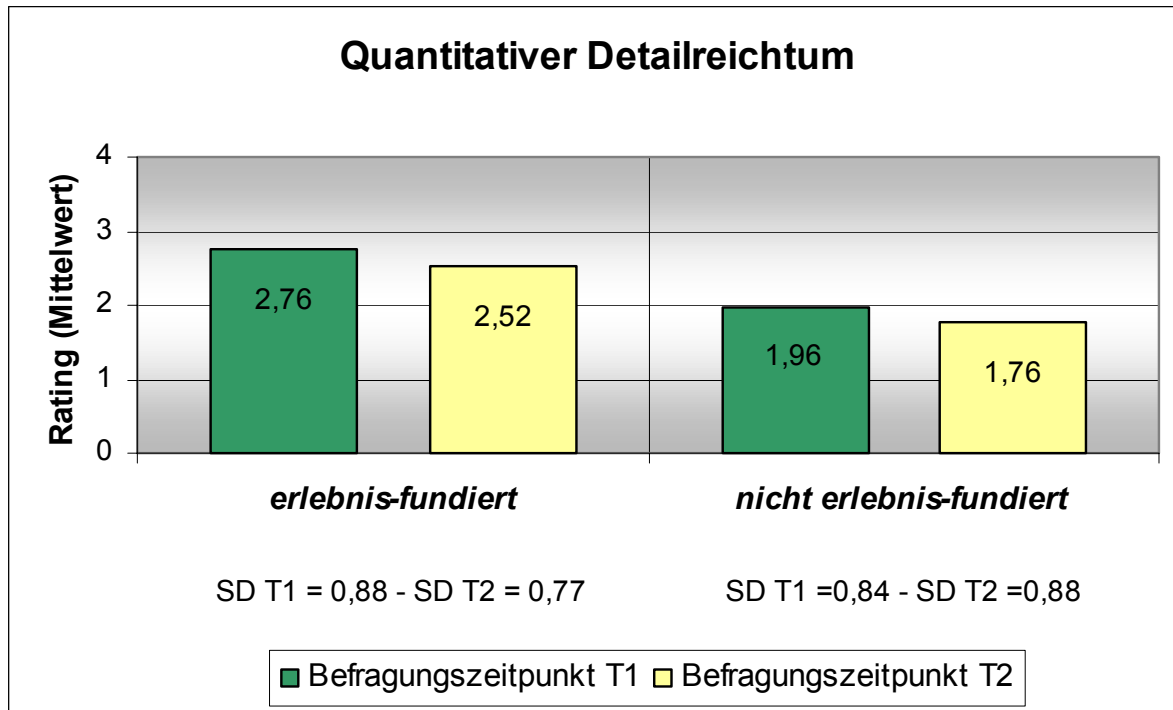
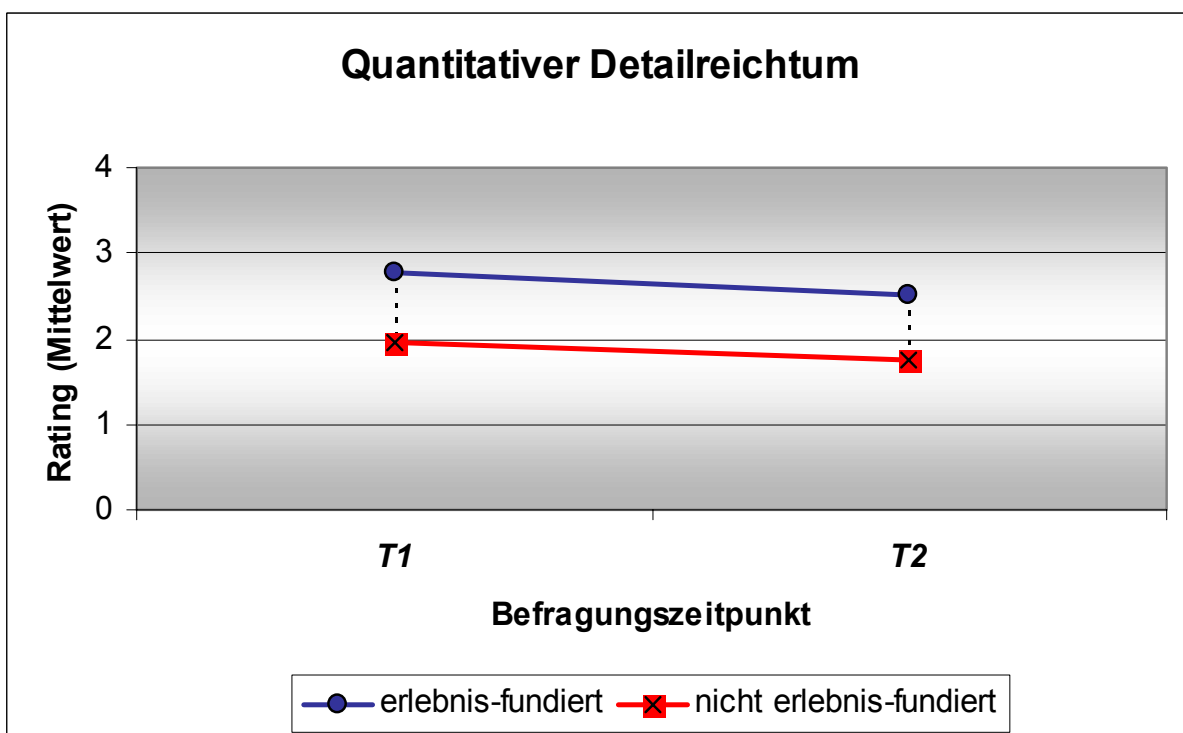


Abbildung 4-26-2: Haupt- und Interaktionseffekte zu T1 und T2: Quantitativer Detailreichtum



Wiederum analoge Ergebnisse zu den beiden ersten Merkmalen zeigen sich auch für den quantitativen Detailreichtum, dem dritten Realkennzeichen der Merkmalskategorie 'Allgemeine Merkmale'.

Der Detailreichtum differenziert ebenfalls sehr gut zwischen erlebnis- und nicht erlebnis-fundierten Geschichten. Das ALM weist hierfür mit $F(1;24) = 16,563$ und $p < 0,001$ wiederum ein höchstsignifikantes Ergebnis aus.

Weiterhin konnten jedoch hinsichtlich des Faktors 'Befragungszeitpunkt' ($F(1;24) = 3,212$ und $p = 0,086$) sowie einer 'Interaktion' ($F(1;24) = 0,026$ und $p = 0,873$) keinerlei statistisch signifikante Befunde ermittelt werden.

Entsprechend dazu ergab der t-Test in der weiteren post-hoc-Analyse für beide Befragungszeitpunkte hoch- bzw. höchstsignifikante Ergebnisse in Form von deutlich geringeren Mittelwerten der erfundenen Geschichten im Vergleich zu den erlebten ($t(24) = 3,266$ und $p = 0,002$ bzw. $t(24) = 3,612$ und $p < 0,001$).

Für beide Geschichten ist im Mittel zwar jeweils ein leichter Abfall von T1 nach T2 in der Merkmalsausprägung zu verzeichnen, jedoch sind diese statistisch unbedeutend, was die post-hoc-Analyse mit $t(24) = 1,445$ und $p = 0,081$ bzw. $t(24) = 1,095$ und $p = 0,142$ bestätigt.

4.2.2.4 Kontextuelle Einbettung

Abbildung 4-27-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Kontextuelle Einbettung

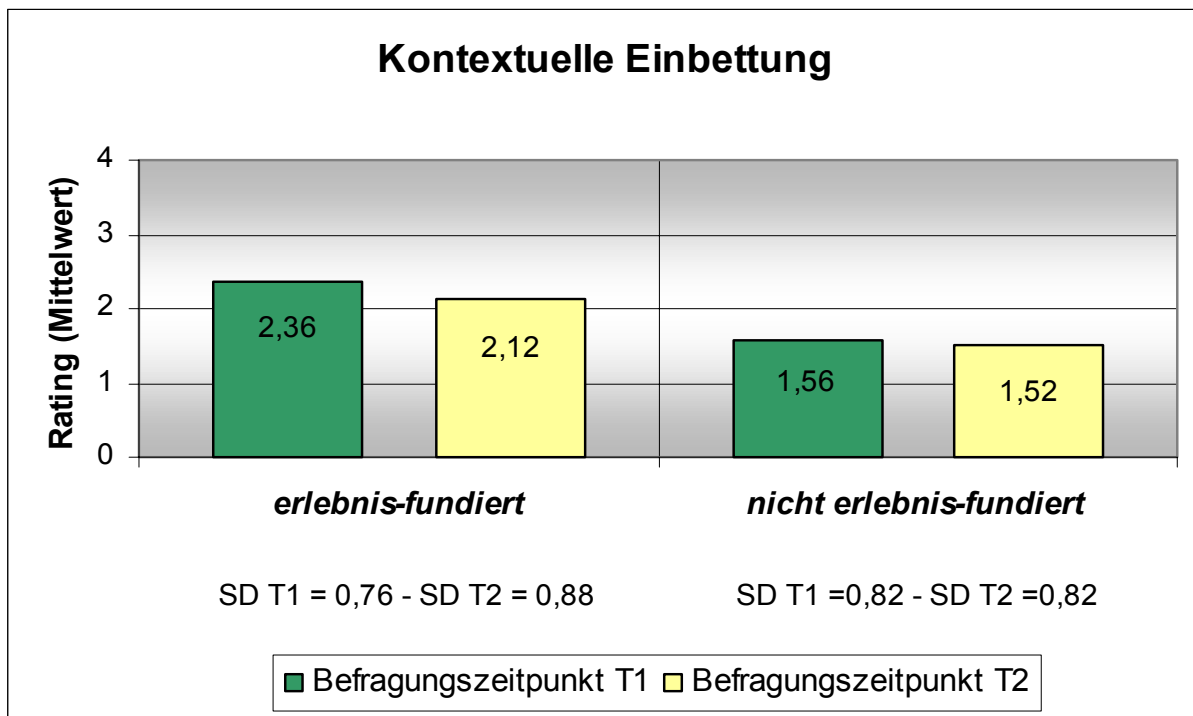
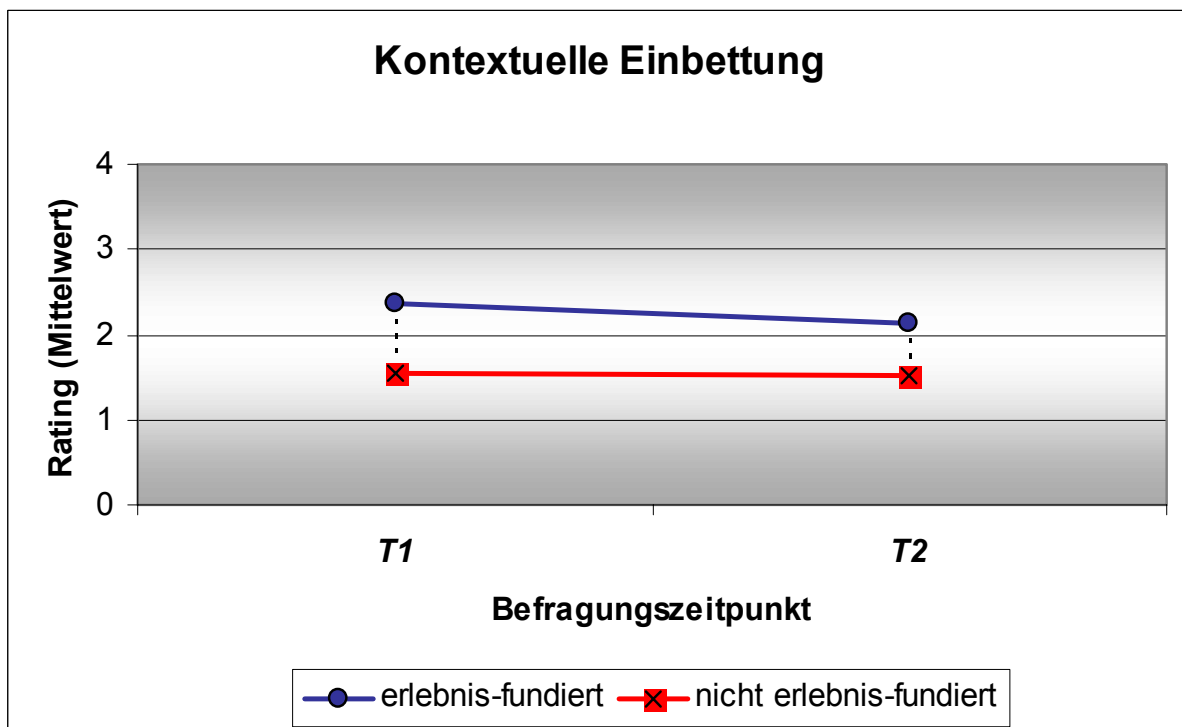


Abbildung 4-27-2: Haupt- und Interaktionseffekte zu T1 und T2: Kontextuelle Einbettung



Auch im Bereich der kontextuellen Einbettung wiederholt sich das schon bekannte Ergebnis. Die univariate ALM-Analyse mit Messwiederholung errechnet auch hier wiederum einen sehr deutlichen Unterschied zwischen erlebnis- und nicht erlebnis-basierten Geschichten, was sich in einem hochsignifikanten Testergebnis zeigt ($F(1;24) = 12,783$ und $p = 0,002$).

Erneut ergeben die Berechnungen bezüglich des Faktors 'Befragungszeitpunkt' sowie einer 'Wechselwirkung' zwischen den Faktoren keine signifikanten Befunde ($F(1;24) = 1,205$ und $p = 0,283$ bzw. $F(1;24) = 0,750$ und $p = 0,395$).

Die post-hoc-Verfahren (Wilcoxon-Test und t-Test) ergaben zu beiden Befragungszeitpunkten hochsignifikante Unterschiede zwischen den wahren und den erfundenen Geschichten ($z = -2,860$ und $p = 0,002$ bzw. $t(24) = 2,598$ und $p = 0,008$). Die erfundenen Geschichten wurden diesbezüglich jeweils deutlich geringer geratet als die wahren.

Hinsichtlich der Veränderungen über die Befragungszeitpunkte hinweg konnten allerdings weder für die wahren ($t(24) = 1,445$ und $p = 0,081$) noch für die erfundenen Geschichten ($z = -0,225$ und $p = 0,411$) anhand der weiteren post-hoc-Verfahren signifikante Ergebnisse beobachtet werden.

4.2.2.5 Schilderung nonverbaler Aktionen

Abbildung 4-28-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung nonverbaler Aktionen

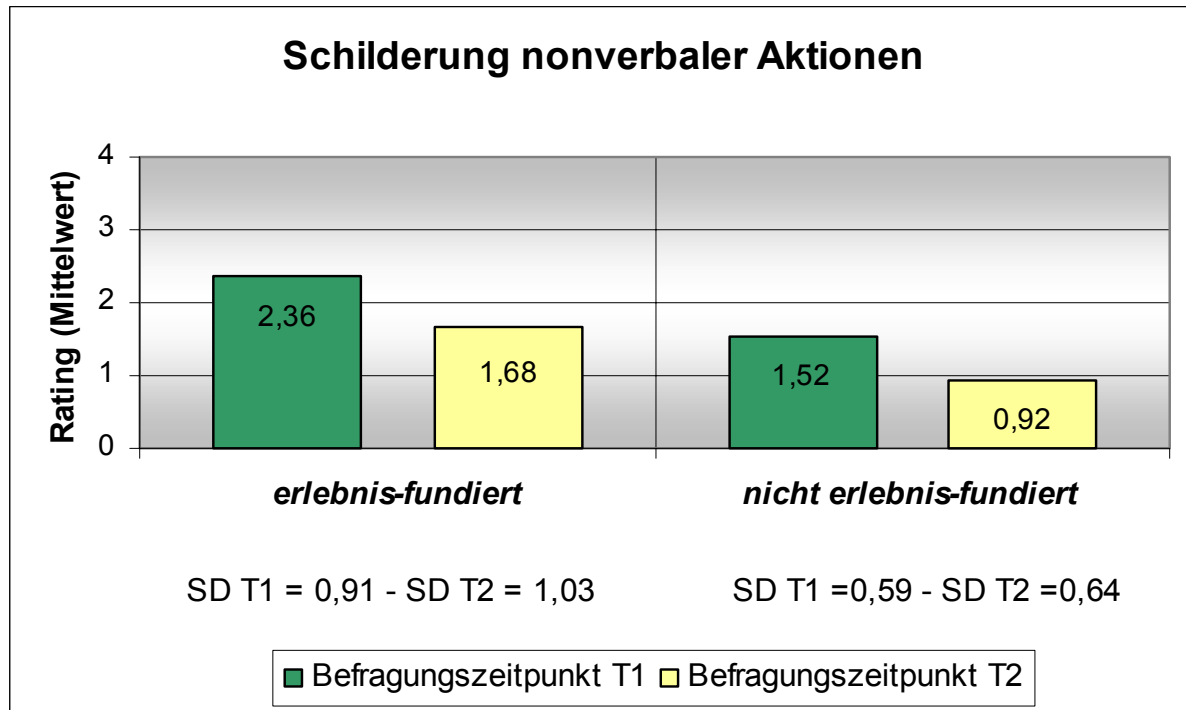
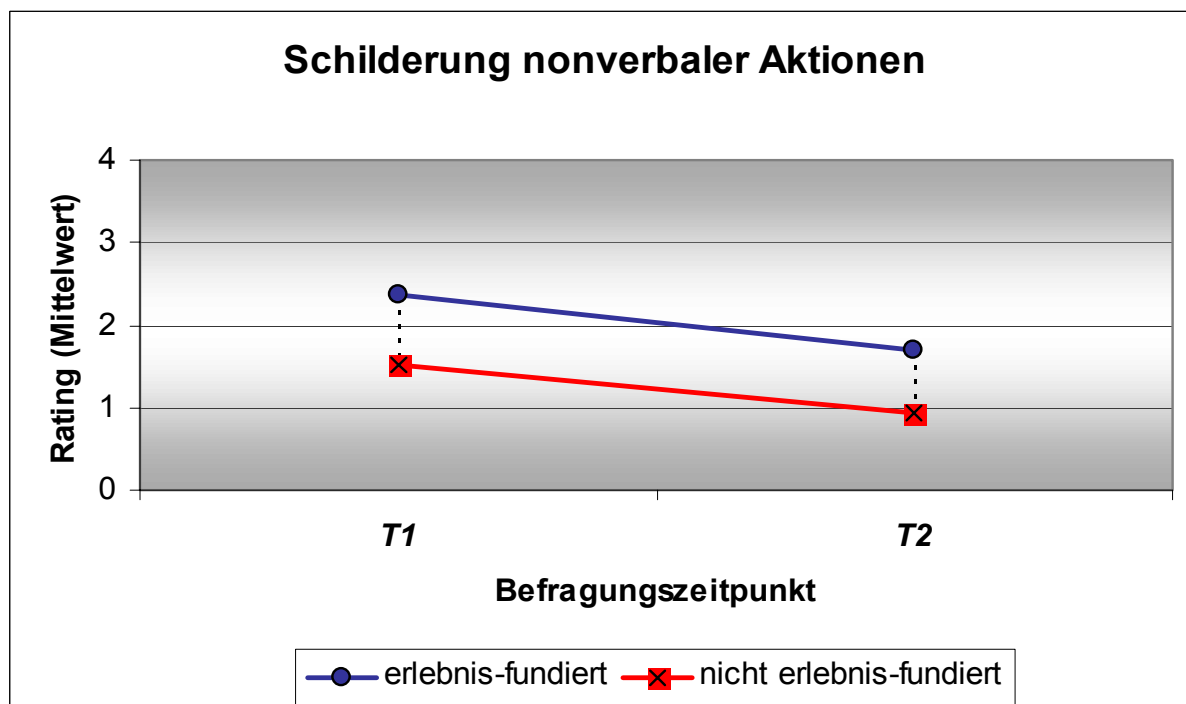


Abbildung 4-28-2: Haupt- und Interaktionseffekte zu T1 und T2: Schilderung nonverbaler Aktionen



Für die Schilderung nonverbaler Aktionen ergeben sich aufgrund des ALM sowohl ein höchstsignifikanter Haupteffekt 'Wahrheitsstatus' ($F(1;24) = 22,588$ und $p < 0,001$) und ebenfalls ein höchstsignifikanter Haupteffekt 'Befragungszeitpunkt' ($F(1;24) = 25,180$ und $p < 0,001$). Eine wechselseitige Abhängigkeit der beiden Faktoren in Form einer Interaktion konnte jedoch auch hier wiederum nicht konstatiert werden ($F(1;24) = 0,069$ und $p = 0,795$).

Bei genauerer post-hoc-Untersuchung anhand des Wilcoxon-Test zeigen sich sowohl zum ersten ($z = -3,080$ und $p = 0,001$) als auch zum zweiten Befragungszeitpunkt ($z = -2,863$ und $p = 0,002$) in den nicht erlebnis-fundierten Aussagen deutlich weniger derartige Schilderungen als in den erlebnis-fundierten.

Analog zum obigen Befund aus dem ALM bestätigen die weiteren post-hoc-Verfahren (Wilcoxon-Test) die wesentlichen Veränderungen der Merkmalsausprägung über die Zeit. Sowohl für die wahren als auch für die erfundenen Geschichten finden sich hochsignifikante Abnahmen von T1 nach T2 im Ausprägungsgrad dieses Realkennzeichens ($z = -2,691$ und $p = 0,004$ bzw. $z = -2,982$ und $p = 0,002$).

4.2.2.6 Wiedergabe von Gesprächen

Abbildung 4-29-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Wiedergabe von Gesprächen

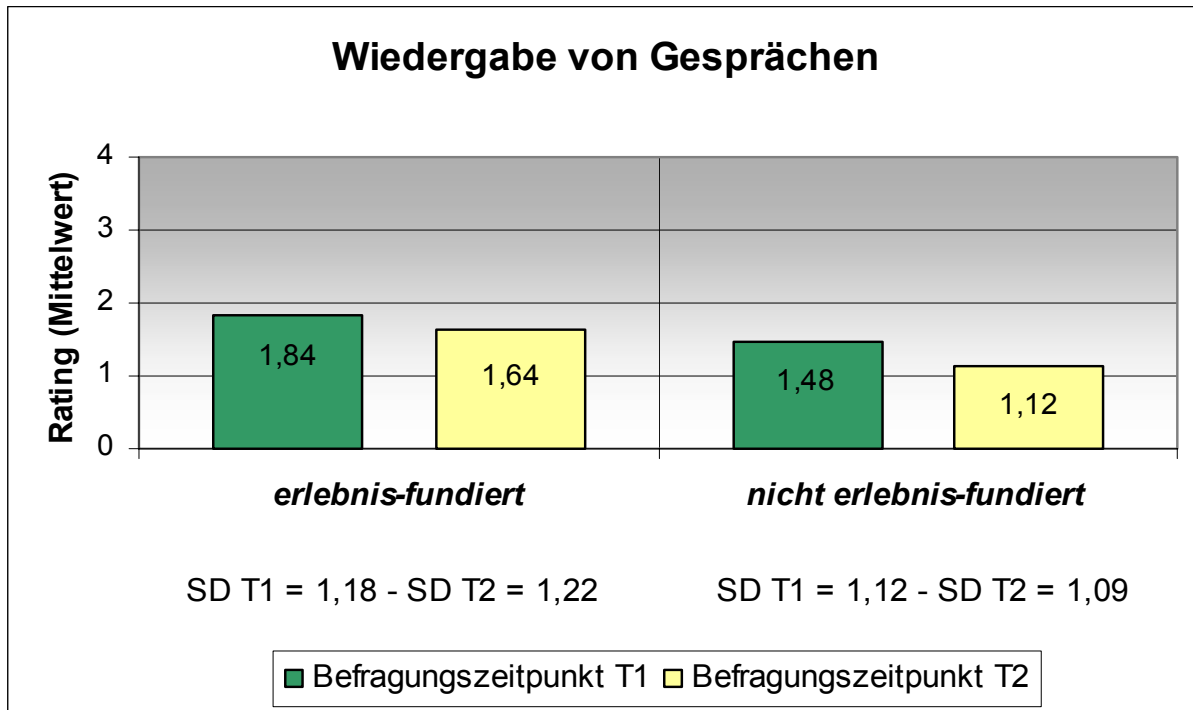
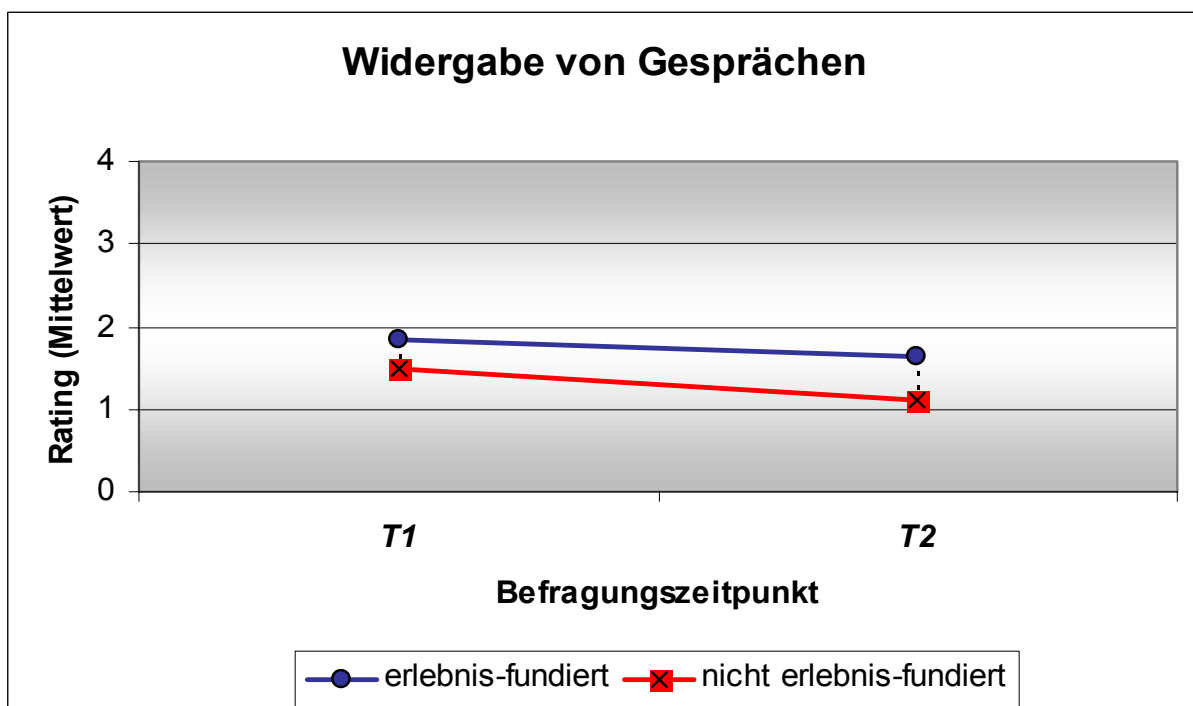


Abbildung 4-29-2: Haupt- und Interaktionseffekte zu T1 und T2: Wiedergabe von Gesprächen



Bezüglich des 'Wahrheitsstatus' ergibt sich hier anhand des univariaten ALM kein signifikanter Haupteffekt ($F(1;24) = 3,169$ und $p = 0,088$), jedoch für den 'Befragungszeitpunkt' ($F(1;24) = 4,931$ und $p = 0,036$). Eine signifikante Interaktion zwischen den Faktoren findet sich mit $F(1;24) = 0,235$ und $p = 0,632$ ebenfalls nicht.

Im weiteren post-hoc-Verfahren anhand des t-Tests zeigt sich jedoch, dass - obwohl zu beiden Zeitpunkten eine Abnahme des Mittelwerts zu beobachten ist - dieses Realkennzeichen zwar zum ersten Befragungszeitpunkt T1 nicht zwischen wahren und erfundenen Geschichten trennt ($t(24) = 1,181$ und $p = 0,125$), jedoch zum zweiten Befragungszeitpunkt T2 schon ($t(24) = 1,797$ und $p = 0,043$).

Betrachtet man die Veränderung über die beiden Befragungszeitpunkte post-hoc genauer, lässt sich feststellen, dass sich nur für die erfundenen Geschichten die Veränderung von T1 nach T2 als signifikant erweist ($t(24) = 1,984$ und $p = 0,030$). Zwar werden innerhalb der erlebten Geschichten zu T2 ebenfalls weniger Gespräche berichtet als zu T1, allerdings erweist sich dieser Unterschied mit $t(24) = 0,866$ und $p = 0,198$ als nicht signifikant.

4.2.2.7 Schilderung von Komplikationen im Handlungsverlauf

Abbildung 4-30-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung von Komplikationen im Handlungsverlauf

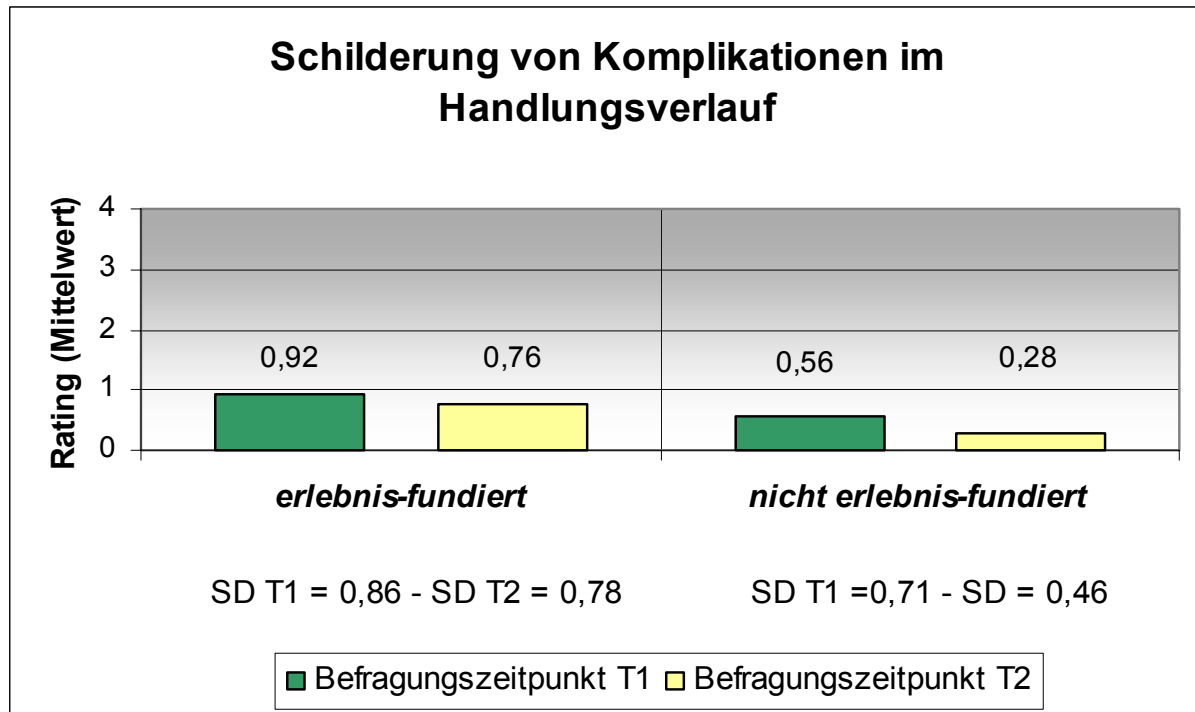
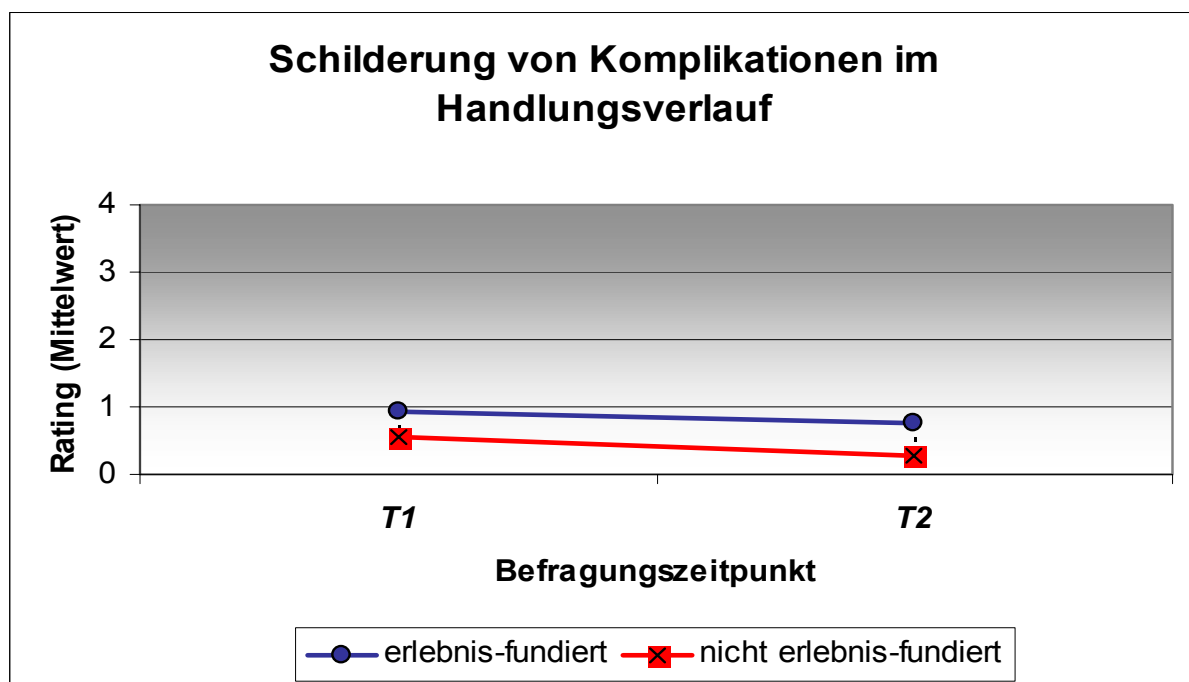


Abbildung 4-30-2: Haupt- und Interaktionseffekte zu T1 und T2: Schilderung von Komplikationen im Handlungsverlauf



Bezüglich berichteter Schilderungen von Komplikationen im Handlungsverlauf ergeben sich sowohl für den Faktor 'Wahrheitsstatus' ($F(1;24) = 5,618$ und $p = 0,026$) als auch für den Faktor 'Befragungszeitpunkt' ($F(1;24) = 5,762$ und $p = 0,024$) signifikante Haupteffekte.

Für eine Wechselwirkung zwischen den beiden Faktoren konnte kein statistisch bedeutsames Ergebnis gefunden werden ($F(1;24) = 0,419$ und $p = 0,524$).

Zwar lassen sich bei genauerer post-hoc-Betrachtung und –Analyse zu beiden Befragungszeitpunkten bei den wahren Geschichten höhere Merkmalsausprägungen beobachten als bei den erfundenen, jedoch erweist sich dieser Unterschied nur zu T2 als statistisch bedeutsam ($z = -2,443$ und $p = 0,008$). Zu T1 konnte kein signifikantes Ergebnis ermittelt werden ($z = -1,485$ und $p = 0,069$).

Ebenso kann bei weiterführender Untersuchung des Effekts 'Befragungszeitpunkt' anhand des Wilcoxon-Test nur für die nicht erlebnis-basierten Geschichten eine signifikante Veränderung ($z = -2,111$ und $p = 0,018$) in Form einer Abnahme der Ausprägung des Merkmals von T1 nach T2 festgestellt werden. Die für die wahren Geschichten auch vorhandene gleichartige Veränderung bleibt statistisch betrachtet allerdings ohne Bedeutung ($z = -1,155$ und $p = 0,124$).

4.2.2.8 Ausgefallene Details

Abbildung 4-31-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Ausgefallene Details

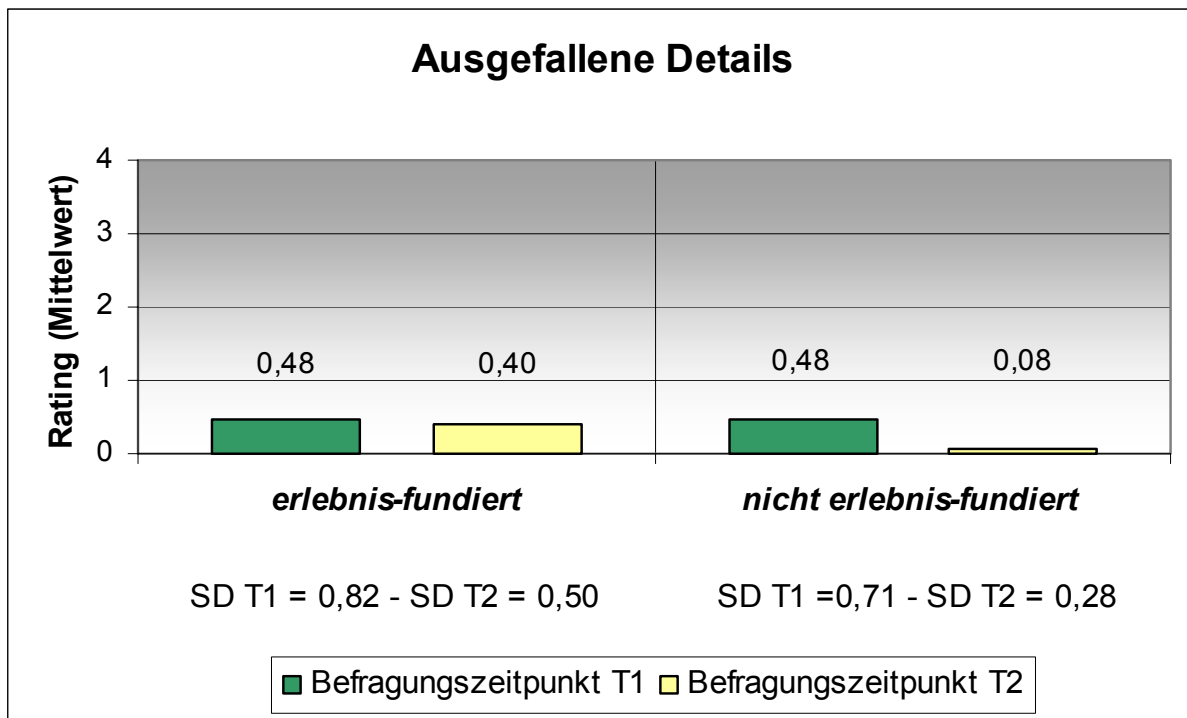
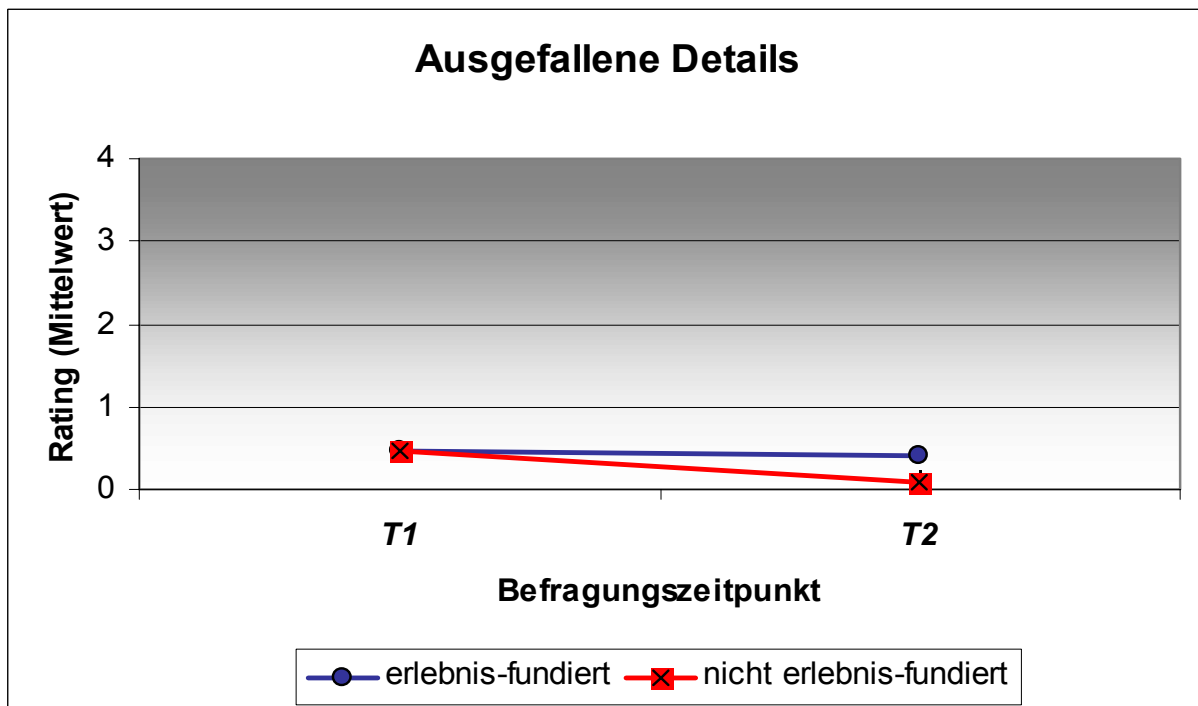


Abbildung 4-31-2: Haupt- und Interaktionseffekte zu T1 und T2: Ausgefallene Details



Das Kriterium 'ausgefallene Einzelheiten' differenziert insgesamt nicht zwischen den wahren und erfundenen Geschichten ($F(1;24) = 1,483$ und $p = 0,235$), jedoch konnte mit $F(1;24) = 5,268$ und $p = 0,031$ ein signifikanter Haupteffekt 'Befragungszeitpunkt' nachgewiesen werden.

Ein signifikantes Testergebnis bezüglich einer gegebenenfalls vorhandenen Interaktion zwischen Wahrheitsstatus der Geschichten und Befragungszeitpunkt konnte ein weiteres Mal nicht beobachtet werden ($F(1;24) = 2,866$ und $p = 0,103$).

Nach Durchführung des Wilcoxon-Test als post-hoc-Verfahren ergab sich jedoch, dass aufgrund der identischen Mittelwerte das Realkennzeichen zwar zum ersten Befragungszeitpunkt nicht zwischen erlebnis-fundierten und nicht erlebnis-fundierten Geschichten trennt ($z = 0,000$ und $p = 0,500$), jedoch in sehr hohem Maß zum zweiten Befragungszeitpunkt ($z = -2,530$ und $p = 0,006$).

Vom ersten zum zweiten Befragungszeitpunkt ergab sich post-hoc für die wahren Geschichten nur eine sehr geringe, nicht signifikante Abnahme in der Ausprägung der Mittelwerte ($z = -0,587$ und $p = 0,279$). Innerhalb der erfundenen Geschichten jedoch gestaltete sich dieser Abfall der Merkmalsausprägung deutlicher und erwies sich im Ergebnis als hochsignifikant ($z = -2,640$ und $p = 0,004$).

4.2.2.9 Nebensächliche Details

Abbildung 4-32-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Nebensächliche Details

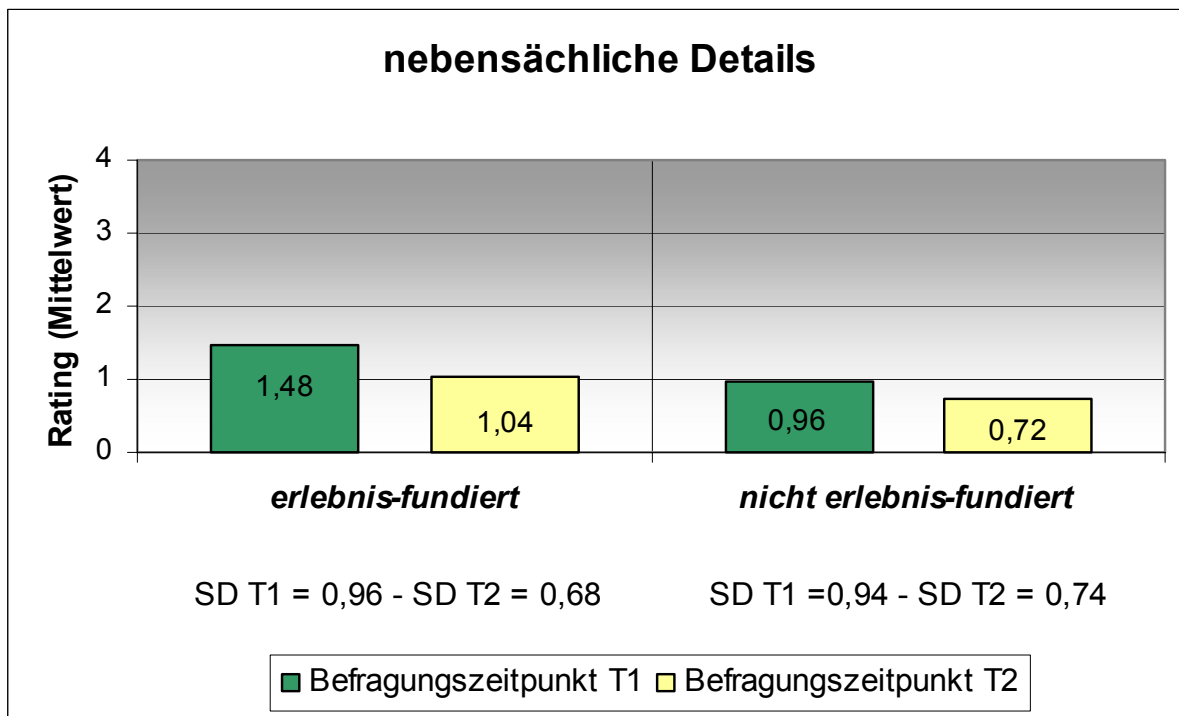
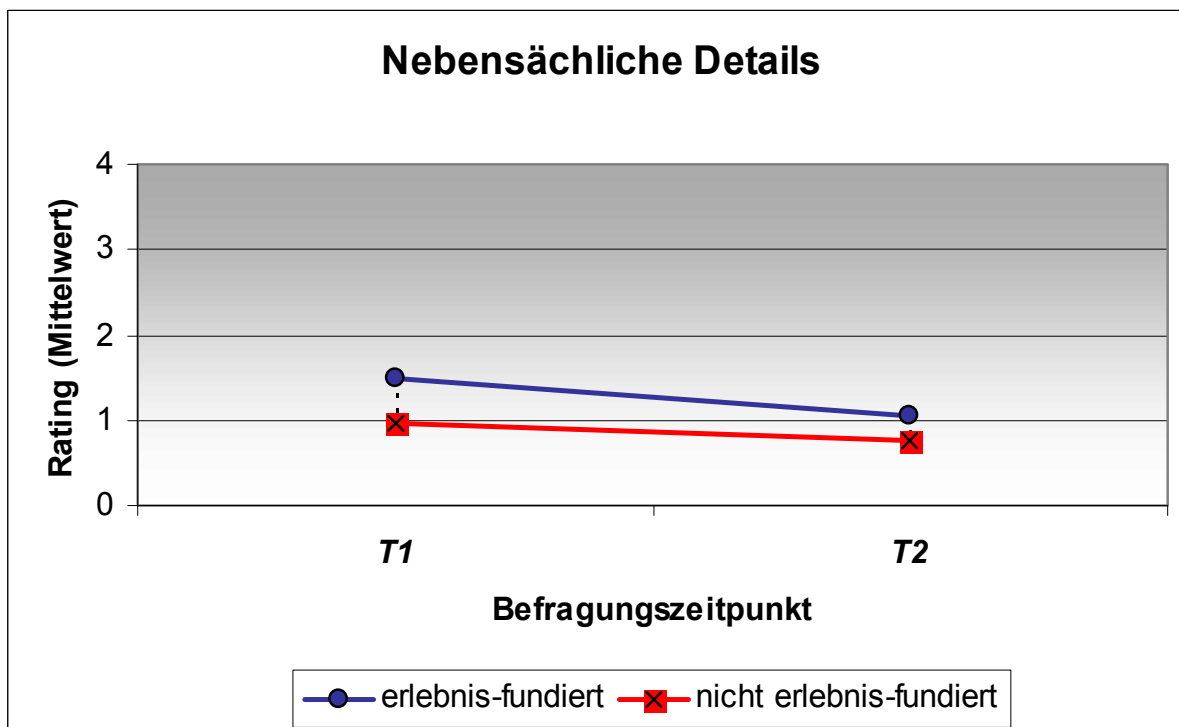


Abbildung 4-32-2: Haupt- und Interaktionseffekte zu T1 und T2: Nebensächliche Details



Die univariate ALM-Analyse ergibt hier für beide Faktoren ('Wahrheitsstatus' und 'Befragungszeitpunkt') signifikante Ergebnisse ($F(1;24) = 6,900$ und $p = 0,015$ bzw. $F(1;24) = 7,034$ und $p = 0,014$).

Eine statistisch relevante Interaktion kann allerdings nicht festgestellt werden ($F(1;24) = 0,444$ und $p = 0,511$).

Die post-hoc-Analyse des Haupteffekts 'Wahrheitsstatus' mittels Wilcoxon-Test bestätigt für beide Befragungszeitpunkte ein in den wahren Geschichten signifikant höheres Ausmaß nebensächlicher Details gegenüber den erfundenen Geschichten ($z = -1,800$ und $p = 0,036$ bzw. $z = -1,795$ und $p = 0,037$).

Bei genauerer Betrachtung der Veränderung der Merkmalsausprägung von T1 nach T2 ergibt die post-hoc-Untersuchung mit $z = -2,230$ und $p = 0,013$ allerdings nur für die erlebnis-fundierten Aussagen eine signifikante Veränderung. Zwar wurden in den erfundenen Geschichten zu T2 ebenfalls weniger nebensächliche Einzelheiten berichtet als zu T1, jedoch erweist sich dieser Unterschied als recht gering und statistisch bedeutungslos ($z = -1,064$ und $p = 0,144$).

4.2.2.10 Schilderung eigener psychischer oder physischer Vorgänge

Abbildung 4-33-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung eigener psychischer oder physischer Vorgänge

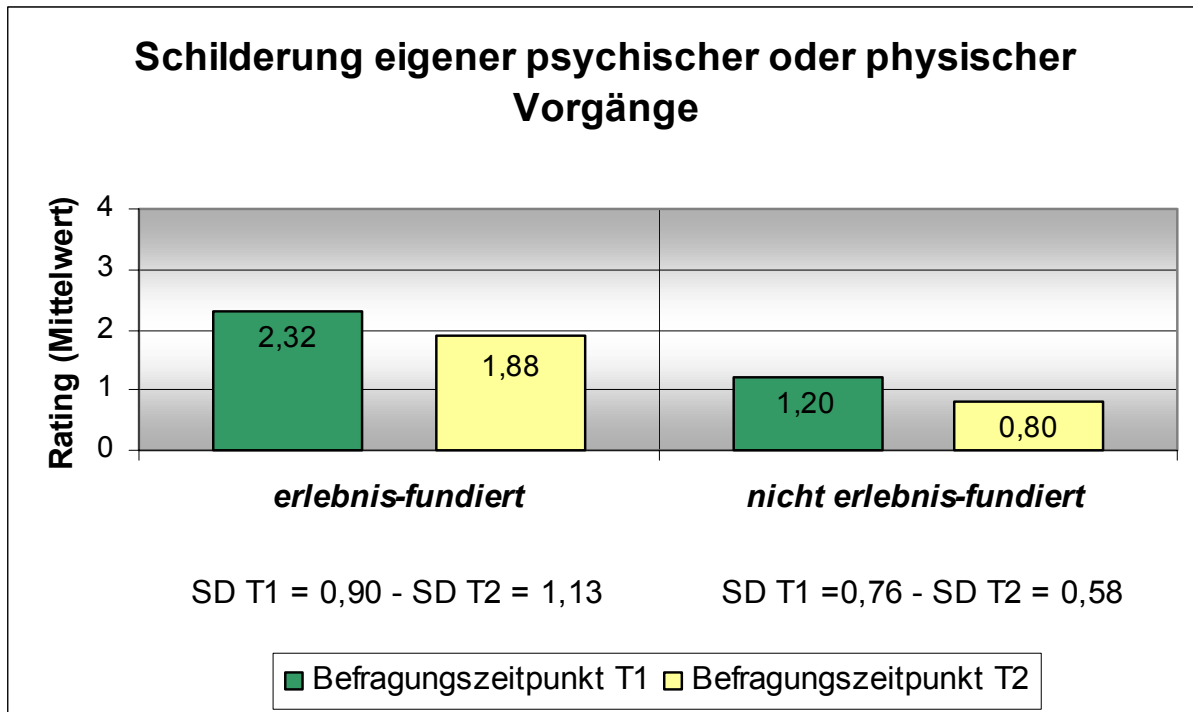
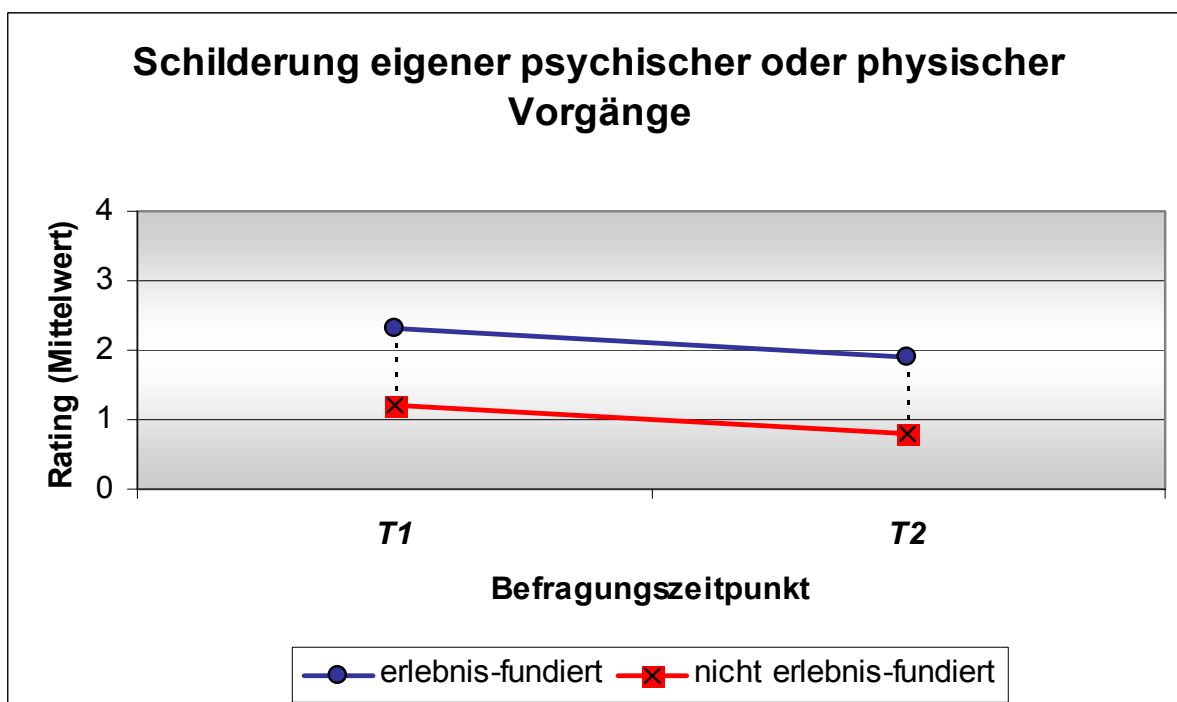


Abbildung 4-33-2: Haupt- und Interaktionseffekte zu T1 und T2: Schilderung eigener psychischer oder physischer Vorgänge



Aufgrund der univariaten ALM-Analyse mit Messwiederholung zeigt sich hier ein höchstsignifikanter Haupteffekt 'Wahrheitsstatus' ($F(1;24) = 27,396$ und $p < 0,001$). Das Realkennzeichen trennt also sehr gut zwischen den Geschichten. Ein ebenfalls deutlicher Unterschied manifestiert sich hinsichtlich des Faktors 'Befragungszeitpunkt', wo sich mit $F(1;24) = 11,332$ und $p = 0,003$ ein hochsignifikantes Ergebnis findet.

Eine wechselseitige Abhängigkeit zwischen den Faktoren war auch bei diesem Glaubhaftigkeitskriterium nicht nachzuweisen ($F(1;24) = 0,025$ und $p = 0,877$).

Bei der näheren Analyse des Effekts 'Status' unter Verwendung des Wilcoxon-Test ist zu beiden Befragungszeitpunkten in den erlebnis-fundierten Geschichten der Ausprägungsgrad dieses Realkennzeichens deutlich und höchstsignifikant höher als in den nicht erlebnis-fundierten ($z = -3,345$ und $p < 0,001$ bzw. $z = -3,371$ und $p < 0,001$).

Im weiterhin post-hoc durchgeführten t-Test erwies sich die Abnahme der Merkmalsausprägung von T1 nach T2 für die wahren Geschichten mit $t(24) = 1,901$ und $p = 0,035$ als signifikant. Für die erfundenen Geschichten wurde anhand des Wilcoxon-Test sogar eine höchstsignifikante Veränderung dieser Art post-hoc ermittelt ($z = -3,162$ und $p = 0,001$).

4.2.2.11 Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten

Abbildung 4-34-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten

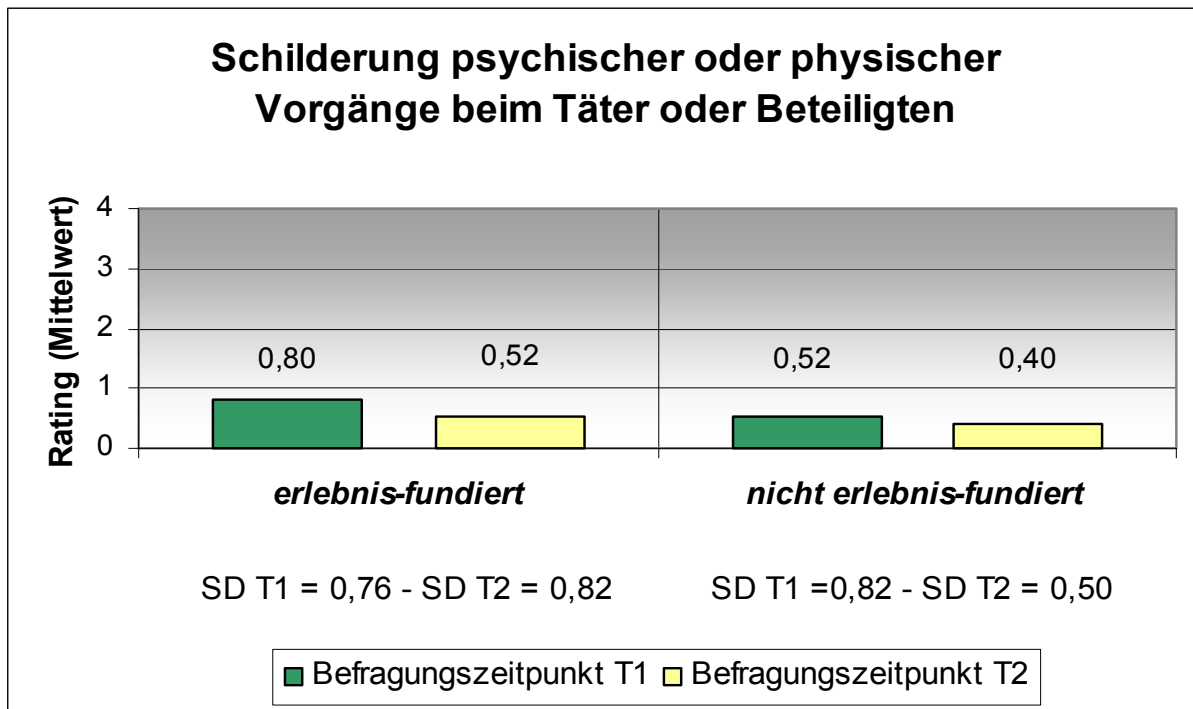
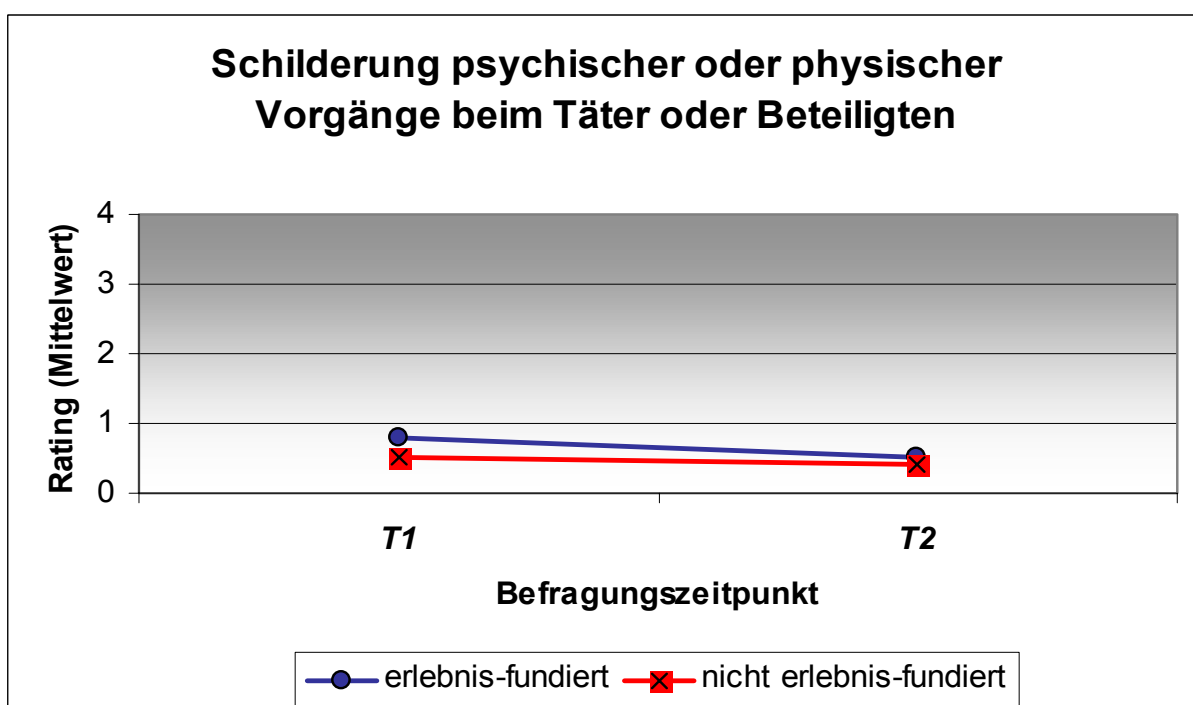


Abbildung 4-34-2: Haupt- und Interaktionseffekte zu T1 und T2: Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten



Für das Realkennzeichen 'Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten' zeigen sich weder für die beiden Faktoren 'Wahrheitsstatus' und 'Befragungszeitpunkt' ($F(1;24) = 1,333$ und $p = 0,260$ bzw. $F(1;24) = 2,667$ und $p = 0,116$) noch für eine wechselseitige Interaktion ($F(1;24) = 0,793$ und $p = 0,382$) signifikante Effekte bzw. Ergebnisse.

Entsprechend gestalten sich auch aufgrund der geringen Mittelwerte und Mittelwertsunterschiede die post-hoc-Ergebnisse des Wilcoxon-Test für Paardifferenzen:

Die Unterschiede in der Merkmalsausprägung zwischen wahren und erfundenen Geschichten sind zu T1 wie auch zu T2 so gering, dass keine signifikanten Ergebnisse ermittelt werden konnten ($z = -1,238$ und $p = 0,109$ bzw. $z = -0,714$ und $p = 0,238$).

Zwar wird das Ergebnis einer signifikanten Veränderung von T1 nach T2 für die erlebnisbasierten Aussagen mit $z = -1,578$ und $p = 0,058$ nur knapp verfehlt, jedoch bleibt dieses auch wie die Veränderung über die Zeit für die nicht erlebnisbasierten Aussagen ($z = -0,905$ und $p = 0,183$) statistisch und im Endergebnis ohne Relevanz. In beiden Fällen kann eine leichte Abnahme vom ersten zum zweiten Befragungszeitpunkt beobachtet werden.

4.2.2.12 Spontane Verbesserungen der eigenen Aussage

Abbildung 4-35-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Spontane Verbesserungen der eigenen Aussage

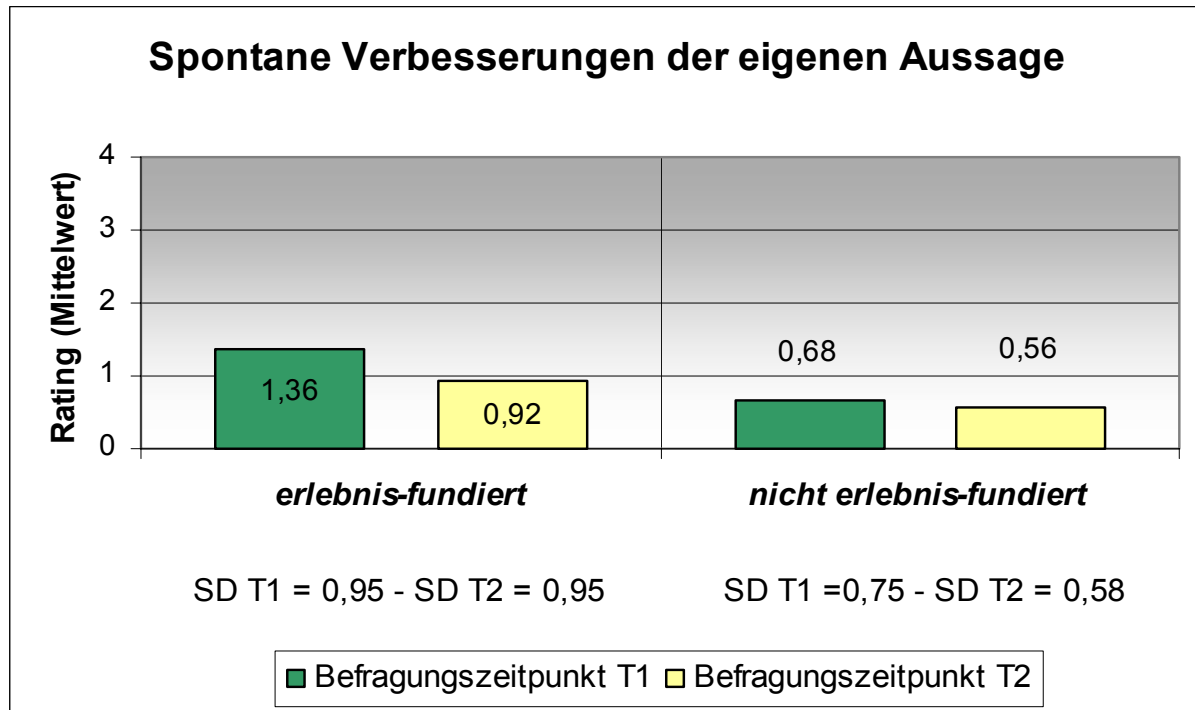
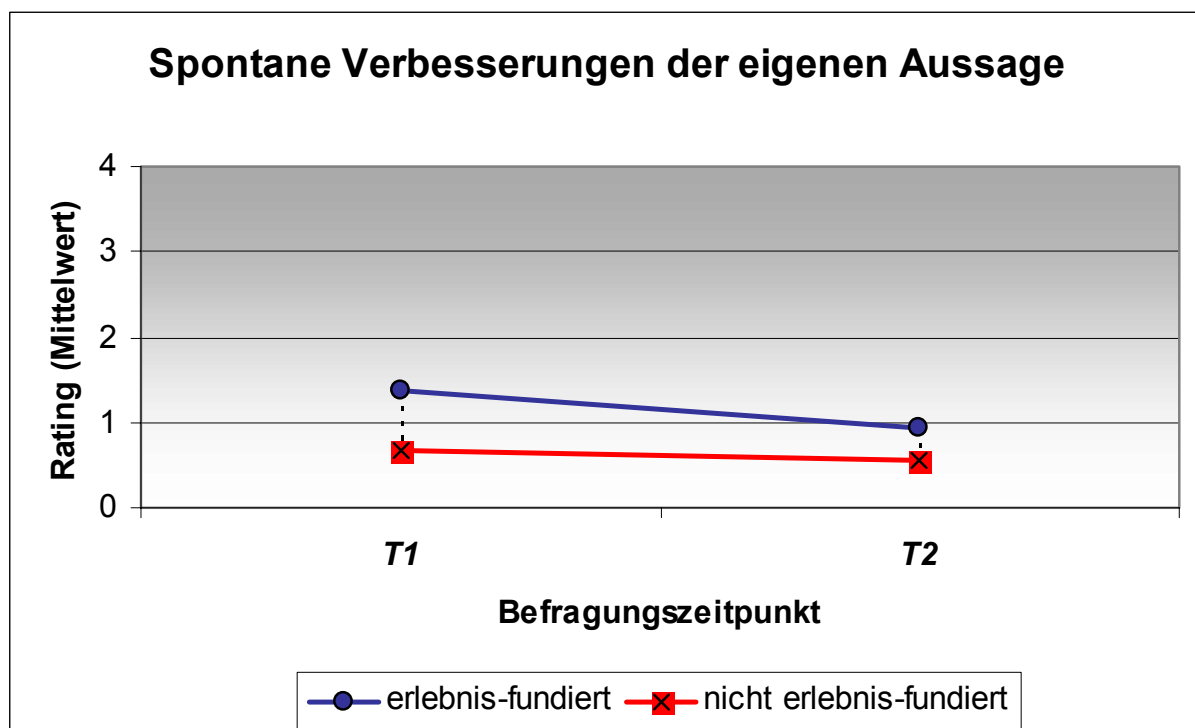


Abbildung 4-35-2: Haupt- und Interaktionseffekte zu T1 und T2: Spontane Verbesserungen der eigenen Aussage



Insgesamt ergibt die varianzanalytische Berechnung mit $F(1;24) = 8,432$ und $p = 0,008$ einen hochsignifikanten Haupteffekt 'Wahrheitsstatus'. Das Kriterium differenziert also sehr gut zwischen erlebnis- und nicht erlebnis-fundierten Aussagen.

Bezogen auf den zeitlichen Aspekt konnte insgesamt betrachtet jedoch kein signifikantes Ergebnis als Haupteffekt ermittelt werden ($F(1;24) = 3,350$ und $p = 0,080$).

Ebenso ist auch keine signifikante Interaktion zwischen 'Wahrheitsstatus' und 'Befragungszeitpunkt' erkennbar ($F(1;24) = 1,641$ und $p = 0,212$).

Die anschließende post-hoc-Analyse bezüglich des hochsignifikanten Haupteffekts 'Status' ließ jedoch erkennen, dass das Realkennzeichen zum ersten Befragungszeitpunkt T1 zwar ebenso hochsignifikant ($z = -2,691$ und $p = 0,004$) zwischen den Geschichten trennt, allerdings nicht zum zweiten Befragungszeitpunkt T2 ($z = -1,639$ und $p = 0,051$). Da zu T2 das Signifikanzniveau jedoch nur denkbar knapp verfehlt wird, kann tendenziell dennoch davon ausgegangen werden, dass 'spontane Verbesserungen der eignen Aussage' auch zu einem zweiten Befragungszeitpunkt eine gute Diskriminanzeigenschaft aufweisen.

Zwar zeigt sich bei genauerer post-hoc-Untersuchung anhand des Wilcoxon-Test mit $z = -0,905$ und $p = 0,183$ ebenso keine signifikante Abnahme in der Ausprägung dieses Glaubhaftigkeitsmerkmals von T1 nach T2 innerhalb der erfundenen Geschichten, jedoch schon für die wahren Geschichten. Hier kann mit $z = -1,689$ und $p = 0,046$ ein signifikantes Ergebnis festgestellt werden.

4.2.2.13 Zugaben von Erinnerungslücken

Abbildung 4-36-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Zugaben von Erinnerungslücken

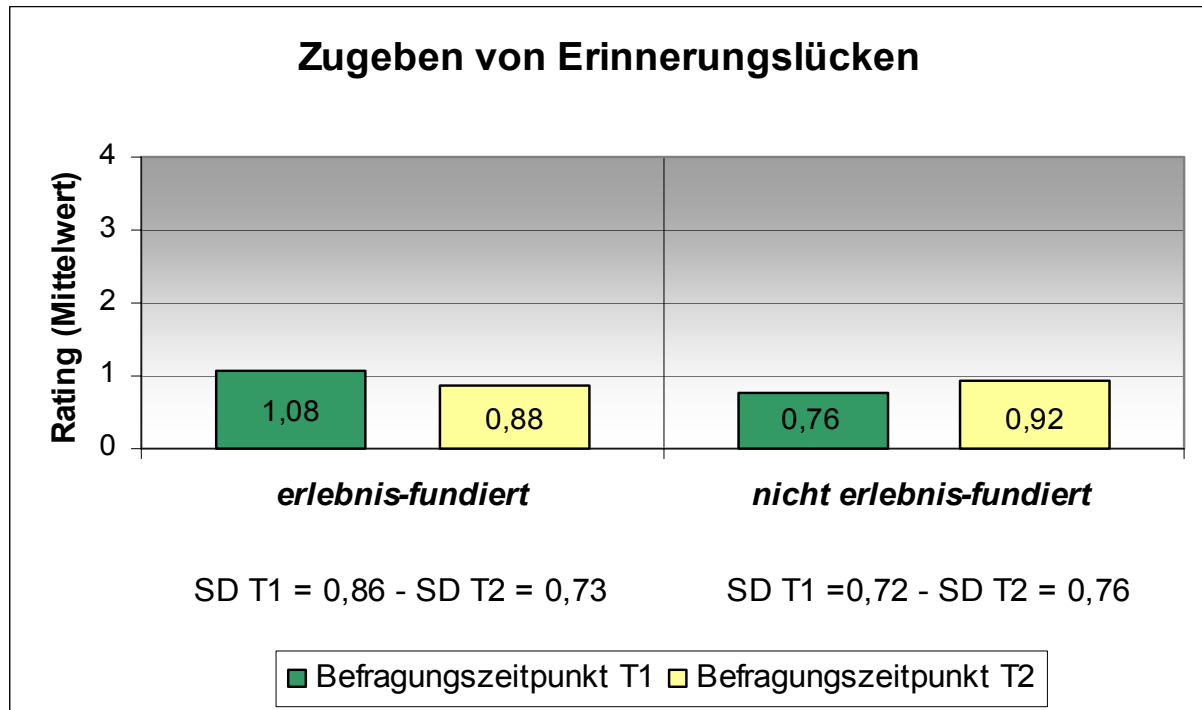
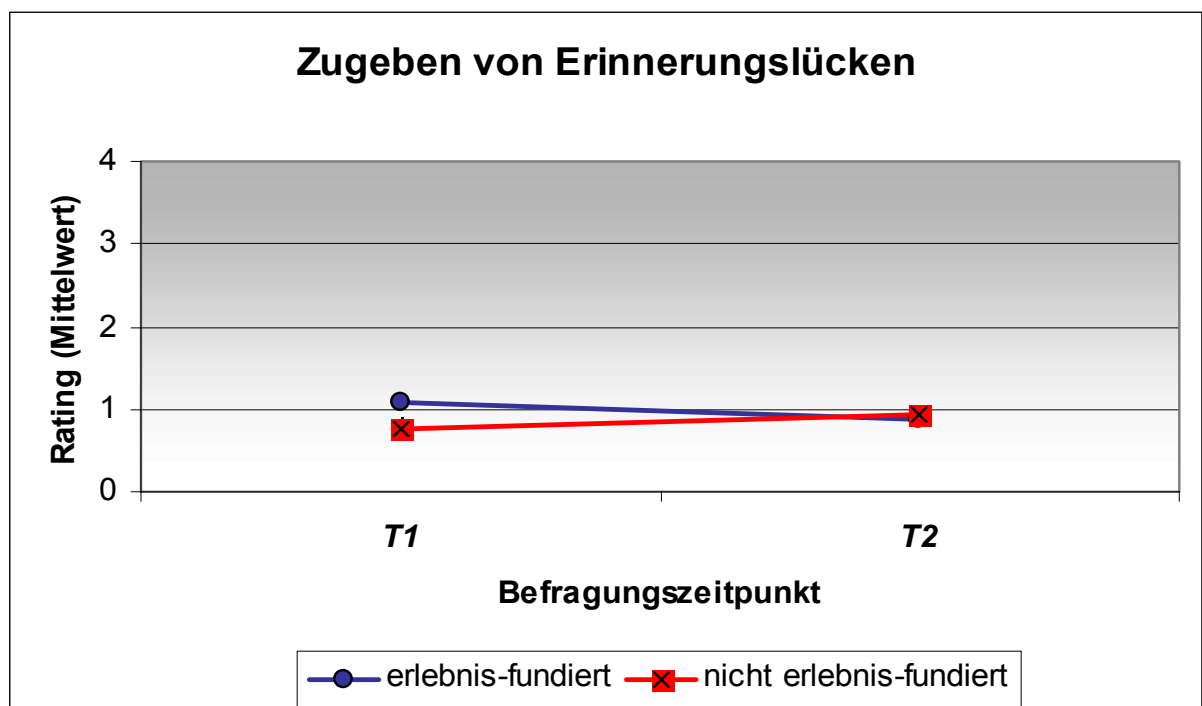


Abbildung 4-36-2: Haupt- und Interaktionseffekte zu T1 und T2: Zugaben von Erinnerungslücken



Das ALM ergibt hinsichtlich des Zugebens von Erinnerungslücken weder für die beiden Faktoren 'Wahrheitsstatus' ($F(1;24) = 0,825$ und $p = 0,373$) und 'Befragungszeitpunkt' ($F(1;24) = 0,025$ und $p = 0,877$) noch für eine Wechselwirkung untereinander ($F(1;24) = 1,628$ und $p = 0,214$) signifikante Haupteffekte bzw. Ergebnisse.

Das post-hoc-Verfahren bestätigt die oben genannten Ergebnisse und zeigt weder zum ersten ($z = -1,342$ und $p = 0,090$) noch zum zweiten Befragungszeitpunkt ($z = -0,232$ und $p = 0,409$) signifikante Unterschiede zwischen erlebnis-basierten Geschichten und Geschichten ohne Erlebnishintergrund. Tendenziell ist hier zu T2 erstmals der Mittelwert der erfundenen Geschichten höher als der der wahren.

Analoge Ergebnisse bringt post-hoc auch die Untersuchung des Faktors 'Befragungszeitpunkt'. Auch hier können sowohl für wahre ($z = -1,221$ und $p = 0,111$) als auch erfundene Geschichten keine signifikanten Veränderungen zwischen den beiden Befragungszeitpunkten beobachtet werden. Zwar ist bezüglich der nicht erlebnis-fundierten Aussagen ein leichter Anstieg der Merkmalsausprägung von T1 nach T2 ersichtlich, jedoch ohne statistische Relevanz ($z = -0,786$ und $p = 0,216$).

4.2.2.14 Einwände gegen die Richtigkeit der eigenen Aussage

Abbildung 4-37-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Einwände gegen die Richtigkeit der eigenen Aussage

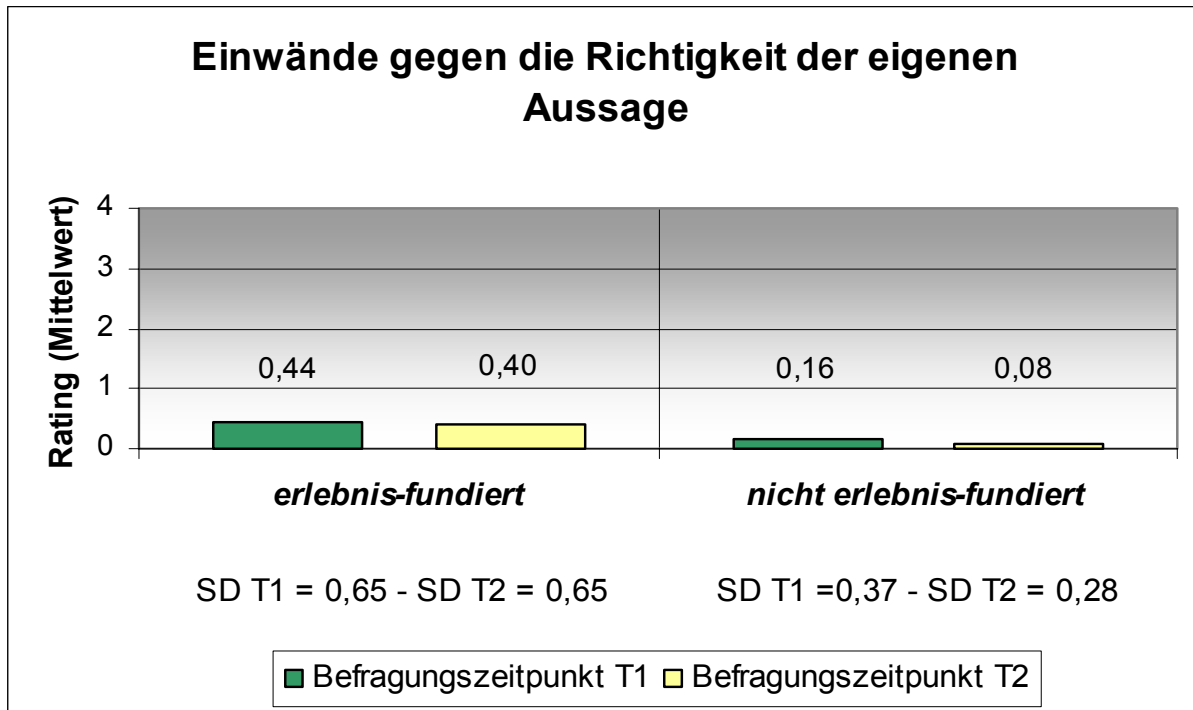
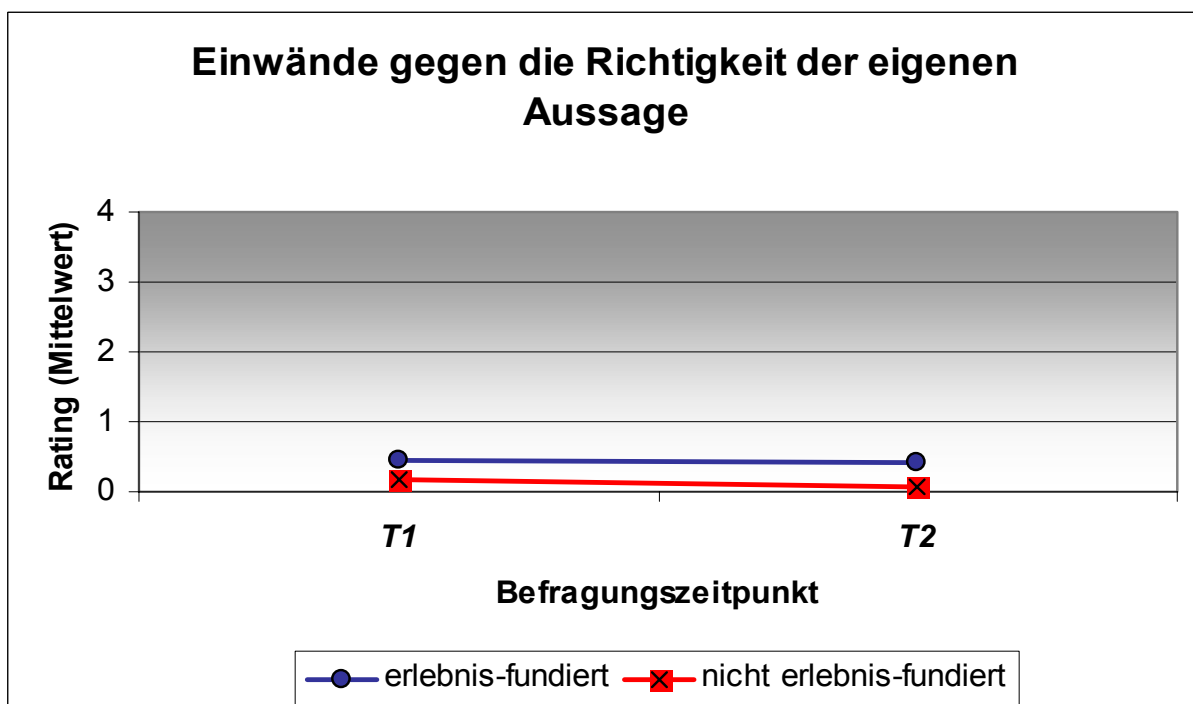


Abbildung 4-37-2: Haupt- und Interaktionseffekte zu T1 und T2: Einwände gegen die Richtigkeit der eigenen Aussage



Für den Faktor 'Wahrheitsstatus' konnte mit $F(1;24) = 7,200$ und $p = 0,013$ ein signifikanter Haupteffekt gefunden werden, jedoch fanden sich keine signifikanten Ergebnisse bezüglich des Faktors 'Befragungszeitpunkt' ($F(1;24) = 0,464$ und $p = 0,503$) sowie einer Wechselwirkung zwischen beiden ($F(1;24) = 0,036$ und $p = 0,852$).

Anhand einer mittels des Wilcoxon-Test durchgeführten post-hoc-Analyse konnte festgestellt werden, dass zu beiden Befragungszeitpunkten in den nicht erlebnis-fundierten Geschichten signifikant weniger Einwände gegen die Richtigkeit der eigenen Aussage vorgebracht wurden als in den erlebnis-fundierten (zu T1: $z = -1,706$ und $p = 0,044$ bzw. zu T2: $z = -1,999$ und $p = 0,023$).

Hinsichtlich der Veränderungen über die Befragungszeitpunkte hinweg konnten hierfür allerdings weder für die wahren noch für die erfundenen Geschichten signifikante Befunde im Rahmen der post-hoc-Überprüfung nachgewiesen werden ($z = -0,237$ und $p = 0,407$ bzw. $z = -0,816$ und $p = 0,207$), da die entsprechenden Unterschiede sehr gering sind.

4.2.2.15 Selbstbelastungen

Abbildung 4-38-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Selbstbelastungen

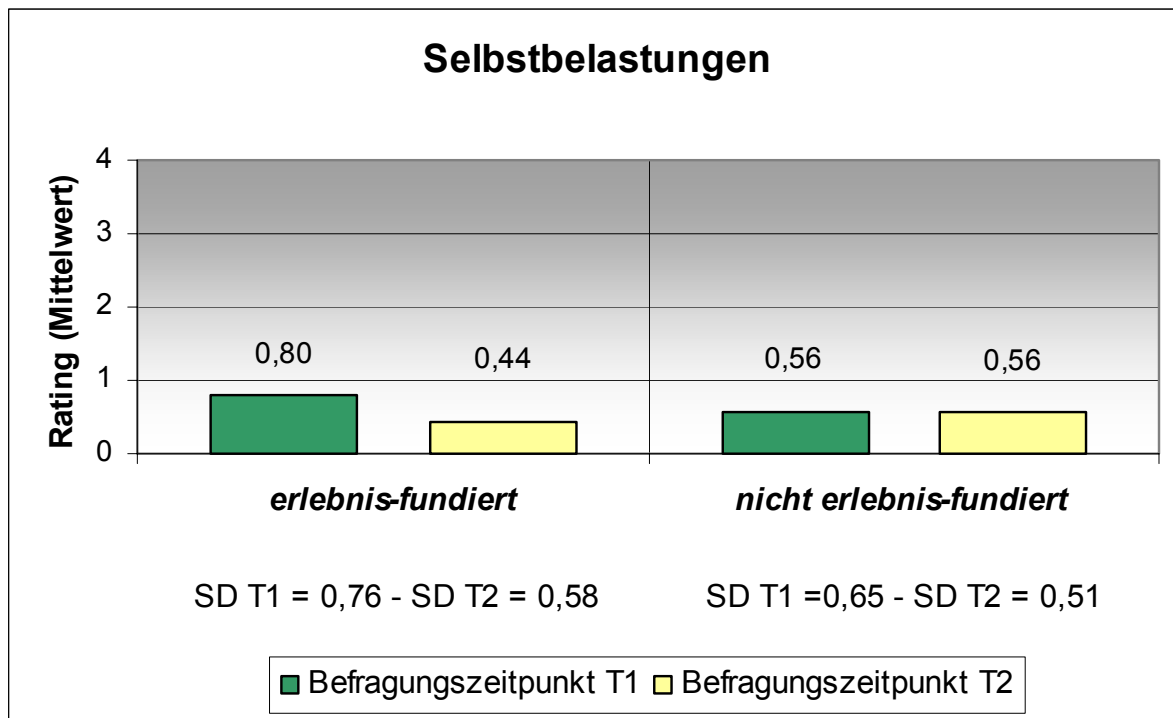
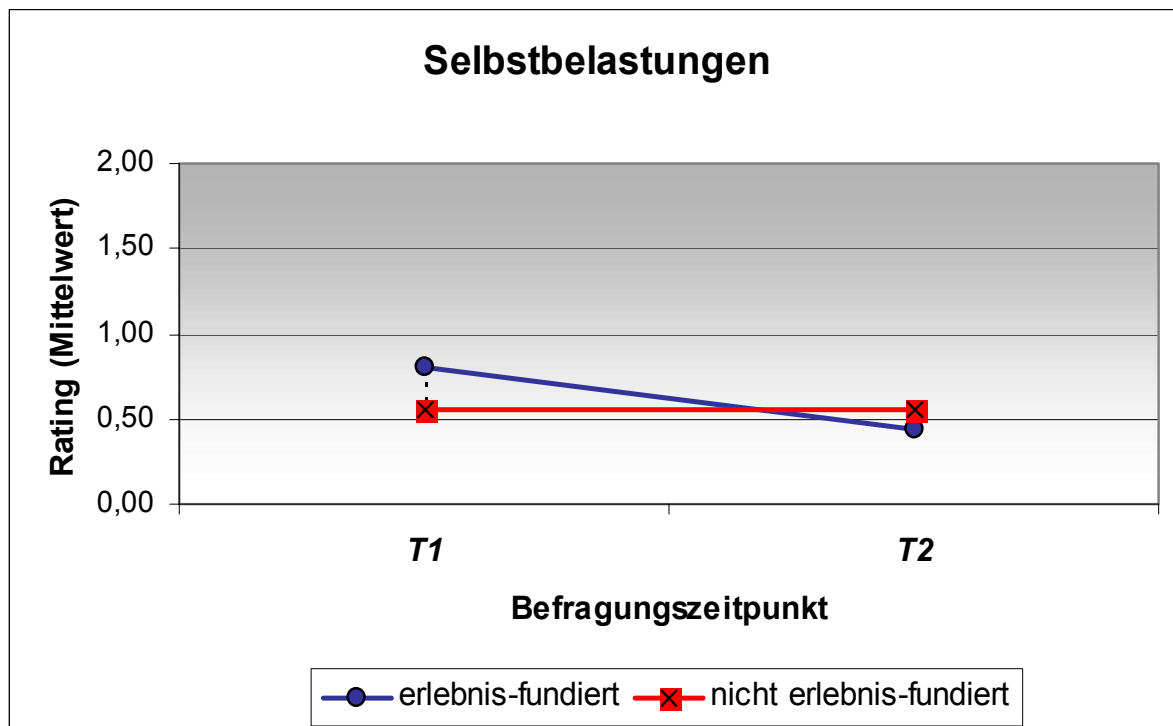


Abbildung 4-38-2: Haupt- und Interaktionseffekte zu T1 und T2: Selbstbelastungen



Im Rahmen der Analyse von vorgebrachten Selbstbelastungen konnten anhand des ALM weder für den 'Wahrheitsstatus' ($F(1;24) = 0,147$ und $p = 0,704$) noch für den 'Befragungszeitpunkt' ($F(1;24) = 2,448$ und $p = 0,131$) signifikante Haupteffekte festgestellt werden.

Des Weiteren zeigte sich ebenfalls kein signifikantes Ergebnis bezüglich einer Interaktion zwischen den beiden Faktoren, obwohl das diesbezügliche Signifikanzniveau mit $F(1;24) = 3,935$ und $p = 0,059$ hier nur knapp verfehlt wird.

Im weiterführenden post-hoc-Verfahren (Wilcoxon-Test) ließ sich zu keinem der beiden Befragungszeitpunkte ein statistisch bedeutsamer Unterschied zwischen den Aussagen mit und denjenigen ohne Erlebnishintergrund nachweisen ($z = -1,209$ und $p = 0,114$ bzw. $z = -0,728$ und $p = 0,234$).

Während sich post-hoc für die erfundenen Geschichten mit $z = 0,000$ und $p = 0,500$ keine signifikante Veränderung vom ersten zum zweiten Befragungszeitpunkt zeigt, fällt der Mittelwert für die wahren Geschichten hinsichtlich berichteter selbstbelastender Aussageelemente von T1 nach T2 signifikant ab ($z = -2,070$ und $p = 0,019$).

4.2.2.16 Inschutznahme des Täters

Abbildung 4-39-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Inschutznahme des Täters oder Beteiligten

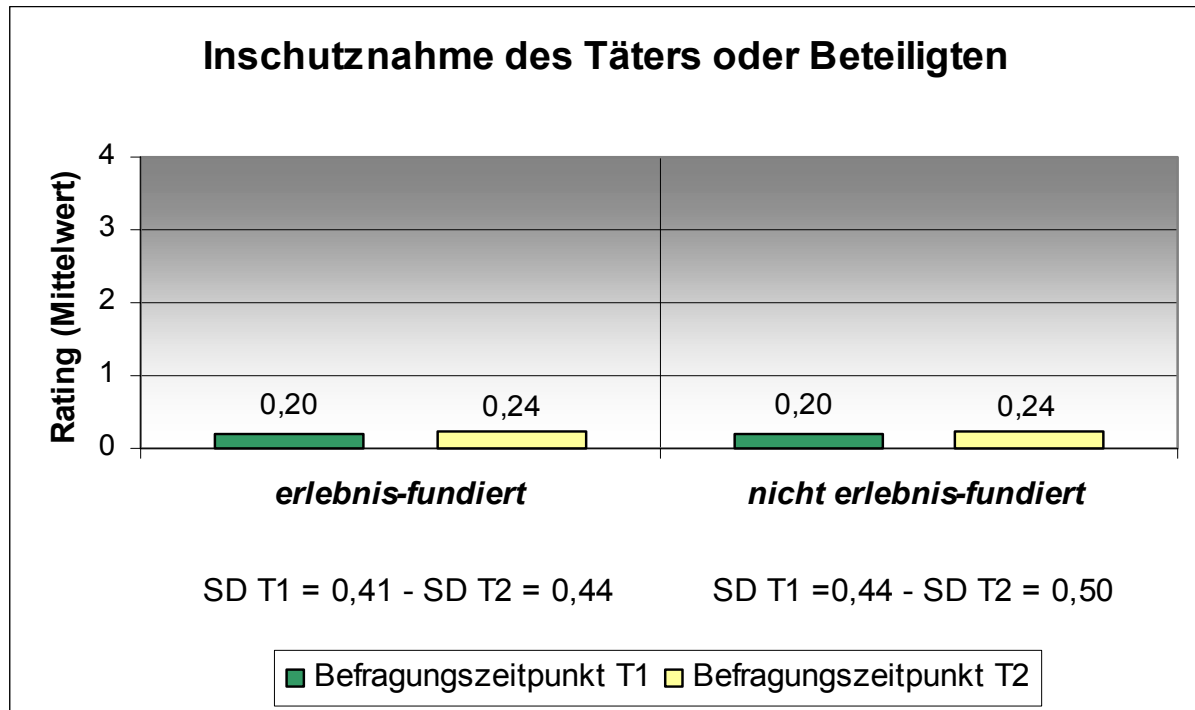
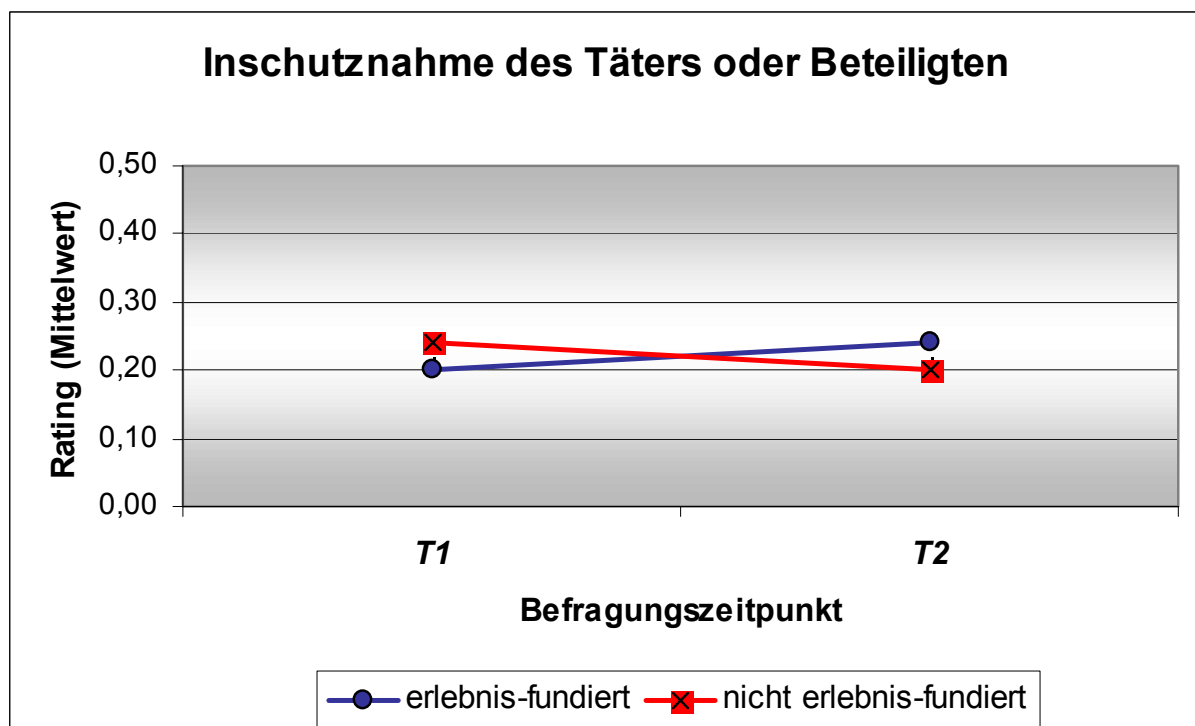


Abbildung 4-39-2: Haupt- und Interaktionseffekte zu T1 und T2: Inschutznahme des Täters oder Beteiligten



Da sich diesbezüglich die Mittelwerte der beiden Geschichten, die Mittelwerte der Befragungszeitpunkte sowie der Veränderungen über die Zeit nicht bzw. fast nicht unterscheiden, können hier keine Haupteffekte (jeweils mit $F(1;24) = 0,000$ und $p = 1,000$) oder Interaktionen ($F(1;24) = 0,242$ und $p = 0,627$) festgestellt werden.

Eine dennoch durchgeführte post-hoc-Untersuchung mittels Wilcoxon-Test ergab weder zu den beiden Befragungszeitpunkten signifikante Unterschiede zwischen wahren und erfundenen Geschichten ($z = -0,302$ und $p = 0,382$ bzw. $z = -0,277$ und $p = 0,391$) noch für wahre oder erfundene Geschichten zeitliche Veränderungen zwischen T1 und T2, die statistisch bedeutsam sind ($z = -0,378$ und $p = 0,353$ bzw. $z = -0,333$ und $p = 0,370$).

4.2.2.17 Deliktspezifische Aussageelemente

Abbildung 4-40-1: Vergleich der Gruppenmittelwerte zu T1 und T2: Deliktspezifische Aussageelemente

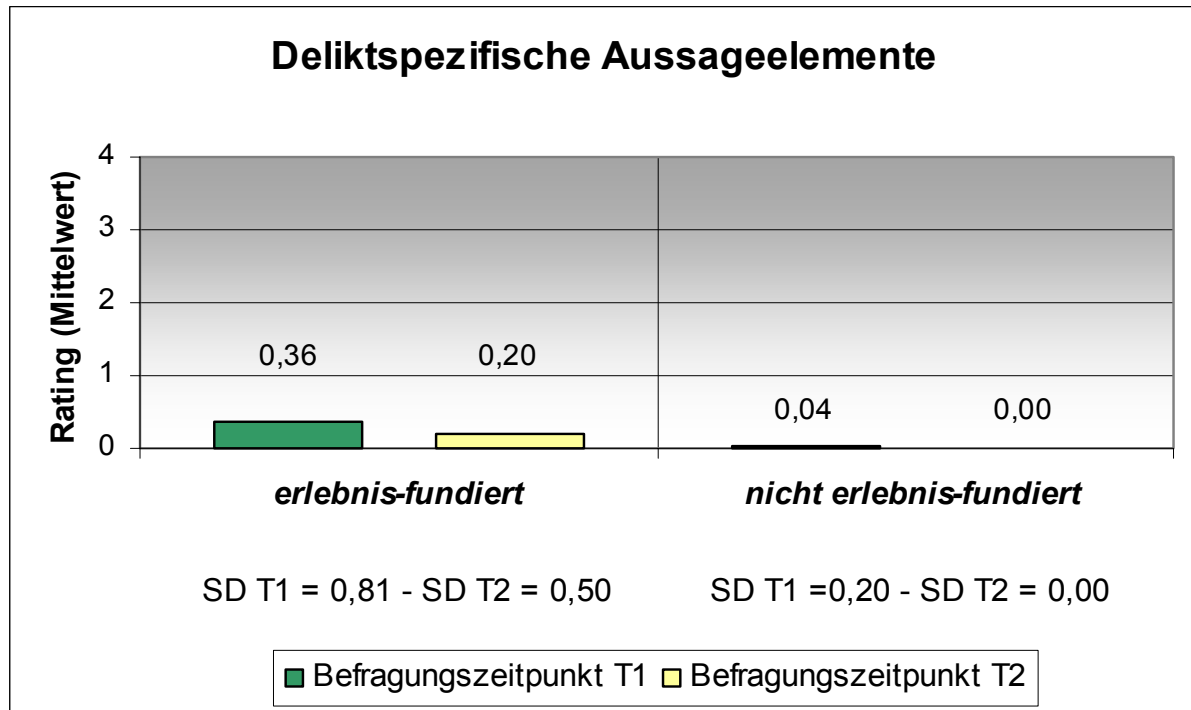
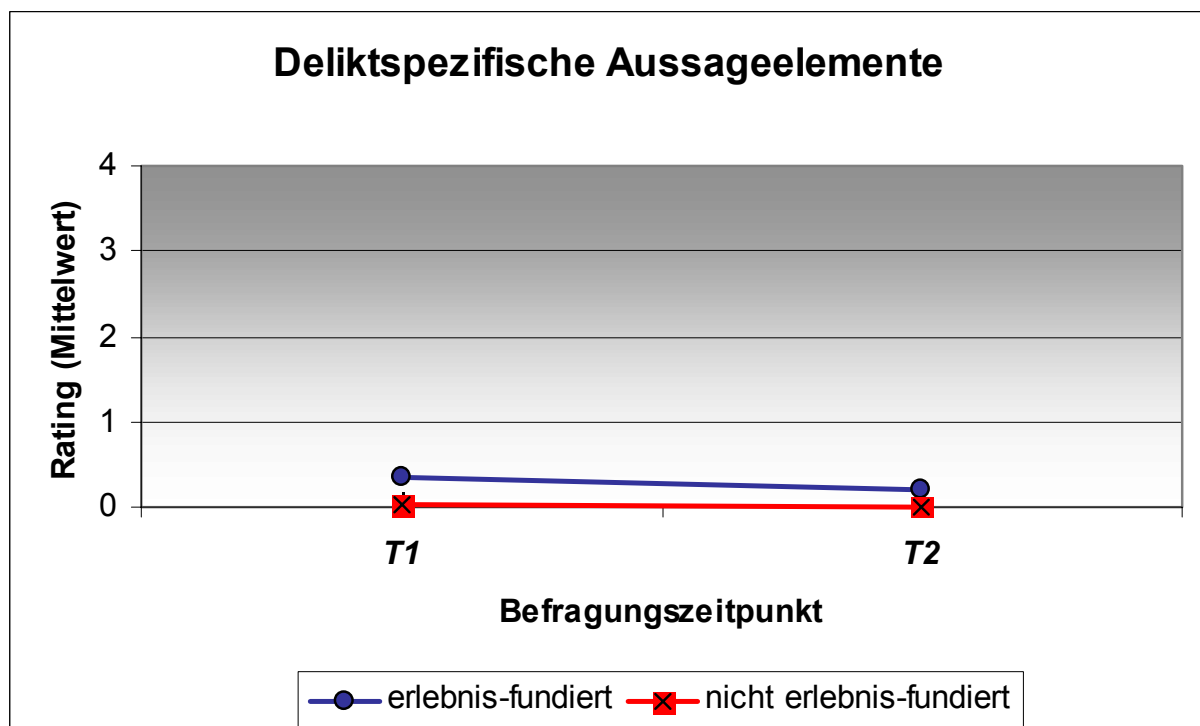


Abbildung 4-40-2: Haupt- und Interaktionseffekte zu T1 und T2: Deliktspezifische Aussageelemente



Zwar bewegen sich die Mittelwerte an sich sowie die Unterschiede hinsichtlich des Vorkommens deliktspezifischer Aussageelemente in einem sehr kleinen Rahmen, dennoch konnte ein (wenn auch gerade noch) signifikanter Haupteffekt 'Wahrheitsstatus' festgestellt werden ($F(1;24) = 4,243$ und $p = 0,050$). Die Aussagen mit Erlebnishintergrund weisen somit signifikant mehr derartige Details bzw. Aussageelemente auf als diejenigen ohne Erlebnisbasis. Hinsichtlich des zweiten Faktors 'Befragungszeitpunkt' sowie einer Interaktion beider Faktoren ergaben sich varianzanalytisch keinerlei statistisch relevante Ergebnisse ($F(1;24) = 3,000$ und $p = 0,096$ bzw. $F(1;24) = 1,000$ und $p = 0,327$).

Der post-hoc berechnete Wilcoxon-Test weist sowohl zu T1 ($z = -1,807$ und $p = 0,036$) als auch zu T2 ($z = -1,890$ und $p = 0,030$) signifikante Ergebnisse aus, dergestalt, dass sich in den erlebnis-basierten Aussagen wesentlich mehr deliktspezifische Aussageelemente finden als in den nicht erlebnis-basierten.

Hinsichtlich einer zeitlichen Veränderung zwischen erstem und zweitem Befragungszeitpunkt konnten jedoch post-hoc weder für wahre noch für erfundene Geschichten signifikante Ergebnisse konstatiert werden ($z = -1,414$ und $p = 0,079$ bzw. $z = -1,000$ und $p = 0,159$).

4.2.3 Ergebnisse zu den Merkmalskategorien und zu allen Aussagemerkmalen als Gesamtinstrument

Da in der wissenschaftlichen Literatur zur Glaubhaftigkeitsbegutachtung sowie in der forensischen Praxis vor allem den ersten drei Realkennzeichen ein besonders hoher diagnostischer Wert zukommt und je nach Abhängigkeit der Frage und den Umständen der Begutachtung auch die übrigen Merkmalskategorien ebenso von speziellem Interesse sein können, wurden diese ebenfalls anhand des ALM sowie post-hoc untersucht.

Ebenso können nach dem schon beschriebenen Prinzip der Aggregation (FIEDLER & SCHMID, 1999; ROSENTHAL, 1987) alle Realkennzeichen zusammengefasst und quasi als Gesamtinstrument zur Differenzierung des Wahrheitsstatus herangezogen werden. Wie ebenfalls schon ausgeführt muss nach DAHLE (1997) allerdings beachtet werden, dass die Glaubhaftigkeitskriterien dabei ungewichtet in die Summenbildung eingehen. Der Vollständigkeit halber wurde jedoch dies ebenfalls statistisch ausgewertet und soll am Ende des Abschnitts auch kurz ausgeführt werden.

4.2.3.1 Allgemeine Merkmale

Für den diagnostisch wichtigen Merkmalskomplex 'Allgemeine Merkmale' konnte anhand des ALM mit $F(1;24) = 46,944$ und $p < 0,001$ ein höchstsignifikanter Haupteffekt 'Wahrheitsstatus' festgestellt werden. Diese erste Kategorie trennt also zunächst unabhängig vom Befragungszeitpunkt sehr deutlich zwischen den Aussagen.

Bezüglich des Faktors 'Befragungszeitpunkt' sowie einer Wechselwirkung zwischen den beiden genannten Faktoren ließen sich jedoch keine signifikanten Ergebnisse finden ($F(1;24) = 3,005$ und $p = 0,096$ bzw. $F(1;24) = 0,667$ und $p = 0,422$).

Der post-hoc durchgeführte t-Test zur genaueren Analyse des Wahrheitsstatus zeigte, dass dieser Merkmalskomplex zu beiden Befragungszeitpunkten T1 und T2 sehr gut zwischen den wahren und den erfundenen Geschichten differenzieren kann, was sich in zwei höchstsignifikanten Testergebnissen widerspiegelt ($t(24) = 6,697$ und $p < 0,001$ bzw. $t(24) = 5,217$ und $p < 0,001$).

Gemäß des Ergebnisses des ALM konnte auch mittels des post-hoc-t-Test weder für die wahren Geschichten ($t(24) = 1,565$ und $p = 0,066$) noch für die erfundenen Geschichten ($t(24) = 0,410$ und $p = 0,343$) eine signifikante Veränderung vom ersten zum zweiten Befragungszeitpunkt nachgewiesen werden.

4.2.3.2 Spezielle Inhalte

Ein ebenfalls höchstsignifikanter Haupteffekt 'Wahrheitsstatus' ergab sich auch für die zweite Merkmalskategorie der speziellen Inhalte ($F(1;24) = 16,008$ und $p = 0,001$).

Sehr deutliche Unterschiede in Form eines hochsignifikanten Ergebnisses ($F(1;24) = 12,760$ und $p = 0,002$) konnten hier auch für den Faktor 'Befragungszeitpunkt' beobachtet werden.

Für die Wechselwirkung zwischen 'Status' und 'Zeitpunkt' weist das ALM allerdings keinen signifikanten Wert aus ($F(1;24) = 0,000$ und $p = 1,000$).

Im anschließend durchgeführten post-hoc-Verfahren zeigte sich wie auch im Bereich der allgemeinen Merkmale, dass die Realkennzeichen der Kategorie 'Spezielle Inhalte' ebenfalls zu beiden Befragungszeitpunkten sehr gut zwischen erlebnis- und nicht erlebnisfundierten Aussagen trennen können. Dies verdeutlichen die hoch- bzw. höchstsignifikanten Ergebnisse der t-Tests ($t(24) = 3,048$ und $p = 0,003$ bzw. $t(24) = 3,718$ und $p < 0,001$).

Ein entsprechender ähnlicher Befund gilt auch für die post-hoc-Analyse des Befragungszeitpunktes. Auch hier fanden sich erwartungsgemäß zu T2 sowohl für die wahren als auch für die erfundenen Geschichten signifikant ($t(24) = 2,138$ und $p = 0,022$) bzw. hochsignifikant ($t(24) = 2,872$ und $p = 0,005$) weniger hohe Werte als zu T1.

4.2.3.3 Inhaltliche Besonderheiten

Auch in diesem Bereich konnten hypothesenkonforme, höchstsignifikante Haupteffekte, jedoch keine signifikante Interaktion ($F(1;24) = 0,024$ und $p = 0,878$) festgestellt werden.

Hinsichtlich des Wahrheitsstatus ergab sich ein F-Wert ($df = 1;24$) von 15,868 sowie ein Signifikanzergebnis von $p = 0,001$, für den Faktor 'Befragungszeitpunkt' $F(1;24) = 18,000$ und $p < 0,001$.

Die Ergebnisse der post-hoc-Untersuchung mittels des t-Tests für abhängige Stichproben konkretisieren den genannten Haupteffekt 'Wahrheitsstatus' für beide Zeitpunkte. Zu T1 wie auch zu T2 sind die Werte für nicht erlebnisfundierte Geschichten im Mittel deutlich geringer als für erlebnisfundierte ($t(24) = 3,186$ und $p = 0,002$ bzw. $t(24) = 3,963$ und $p < 0,001$).

Ein im statistischen Befund sowie ebenso inhaltlich analoges Ergebnis findet sich bei genauerer Betrachtung der Befragungszeitpunkte. Sowohl für die wahren als auch für die erfundenen Geschichten ließ sich ein hoch- bzw. höchstsignifikanter Abstieg der Summe der Merkmalsausprägungen von T1 nach T2 konstatieren ($t(24) = 2,942$ und $p = 0,004$ bzw. $t(24) = 3,412$ und $p = 0,001$).

4.2.3.4 Motivationsbezogene Inhalte

Für die Gruppe der Aussagemerkmale 'Motivationsbezogene Inhalte' konnte mit $F(1;24) = 7,115$ und $p = 0,013$ zwar ein signifikanter Haupteffekt des Faktors 'Wahrheitsstatus' ermittelt werden, jedoch fanden sich weder Hinweise auf einen Haupteffekt 'Befragungszeitpunkt' ($F(1;24) = 2,259$ und $p = 0,146$) noch auf einen statistisch bedeutsamen Interaktionseffekt ($F(1;24) = 1,970$ und $p = 0,173$).

Bei näherer Betrachtung des Statureffekts zeigte sich nach Durchführung von post-hoc-t-Tests, dass die Werte der wahren Geschichte dieser Merkmalskategorie zwar zum ersten Befragungszeitpunkt noch deutlich, d.h. hochsignifikant höher sind als die der erfundenen ($t(24) = 3,177$ und $p = 0,002$), sich dieser Unterschied zum zweiten Befragungszeitpunkt als nicht mehr so deutlich erweist und im Ergebnis auch nicht signifikant ist ($t(24) = 1,039$ und $p = 0,155$).

Die Post-hoc-Analyse des Faktors 'Befragungszeitpunkt' bestätigt zwar insgesamt das Ergebnis des ALM, indem sich weder für die wahren ($t(24) = 1,698$ und $p = 0,051$) noch für die erfundenen Aussagen ($t(24) = 0,225$ und $p = 0,412$) eine signifikante Veränderung von T1 nach T2 zeigt, jedoch wird das Signifikanzniveau hinsichtlich der wahren Geschichten nur denkbar knapp verfehlt. Somit kann wohl davon ausgegangen werden, dass diesbezüglich dieser Merkmalskomplex recht gut und deutlich zwischen den Befragungszeitpunkten differenzieren kann.

4.2.3.5 Deliktspezifische Inhalte

In dieser Untersuchung entspricht diese Kategorie dem einzelnen Aussagemerkmal 'Deliktsspezifische Aussageelemente', soll aber der Vollständigkeit halber ebenso als Kategorie angeführt werden. Für eine graphische Darstellung sei deshalb auf das entsprechende Kapitel der Ausführungen zu den einzelnen RKZ verweisen.

Trotz der geringen Mittelwerte konnte hier dennoch zumindest ein – wenn auch knapp – signifikanter Haupteffekt 'Wahrheitsstatus' beobachtet werden ($F(1;24) = 4,243$ und $p = 0,050$). Ein statistisch bedeutsames Ergebnis bezüglich des Befragungszeitpunktes oder einer Wechselwirkung zwischen beiden Faktoren liegt allerdings nicht vor ($F(1;24) = 3,000$ und $p = 0,096$ bzw. $F(1;24) = 1,000$ und $p = 0,327$).

Der post-hoc durchgeführte Wilcoxon-Test für Paardifferenzen ergab hinsichtlich des Haupteffekts 'Wahrheitsstatus' sowohl für den ersten als für den zweiten Befragungszeitpunkt signifikante Ergebnisse ($z = -1,807$ und $p = 0,036$ bzw. $z = -1,890$ und $p = 0,030$).

Für den Faktor 'Befragungszeitpunkt' konnten anhand des Wilcoxon-Test post-hoc weder für die wahren Geschichten ($z = -1,414$ und $p = 0,079$) noch für die erfundenen ($z = -1,000$ und $p = 0,159$) signifikante Veränderungen zwischen den Befragungszeitpunkten gefunden werden.

4.2.3.6 Aussagemerkmale 'total'

Betrachtet man nun alle in die Auswertung letztendlich miteinbezogenen Aussagemerkmale zusammen (als Summenwert bzw. Gesamtdiagnoseinstrument) ergeben sich folgende Befunde:

Es zeigen sich höchstsignifikante Haupteffekte 'Wahrheitsstatus' und 'Befragungszeitpunkt' ($F(1;24) = 28,821$ und $p < 0,001$ bzw. $F(1;24) = 13,860$ und $p = 0,001$). Die Realkennzeichen sind also insgesamt, zusammengenommen sehr gut geeignet zwischen wahren und erfundenen Geschichten einerseits und zwei Befragungszeitpunkten andererseits zu differenzieren.

Eine wechselseitige Abhängigkeit in Form einer signifikanten Interaktionseffekts liegt allerdings nicht vor ($F(1;24) = 0,556$ und $p = 0,463$).

Bei genauerer post-hoc-Betrachtung des Haupteffekts 'Wahrheitsstatus' bestätigt und konkretisiert sich das oben beschriebene Ergebnis dergestalt, dass sich sowohl zu T1 als auch zu T2 die wahren Geschichten von den erfundenen höchstsignifikant unterscheiden ($t(24) = 4,690$ und $p < 0,001$ bzw. $t(24) = 4,370$ und $p < 0,001$).

Hinsichtlich des Befragungszeitpunktes ergibt die post-hoc-Analyse ebenso für die wahren wie für die erfundenen Geschichten eine deutliche Veränderung von T1 nach T2 in Form eines Abfalls der Merkmalsausprägungen, was sich statistisch in hoch- bzw. signifikanten Testergebnissen widerspiegelt ($t(24) = 2,496$ und $p = 0,010$ bzw. $t(24) = 2,425$ und $p = 0,012$).

4.2.4 Tabellarische Kurzzusammenfassung aller Ergebnisse

4.2.4.1 Ergebnisse zu den einzelnen Aussagemerkmalen

Tabelle 4-69: Ergebnisse des ALM (2 - GLH)

Variable (Aussagemerkmal)	Haupteffekt `WAHRHEITS- STATUS`	Haupteffekt `BEFRAGUNGS- ZEITPUNKT`	Interaktion `STATUS` - `ZEITPUNKT`
Logische Konsistenz	***		
Widerspruchsfreie, ungeordnete Darstellung	***		
Quantitativer Detailreichtum	***		
Kontextuelle Einbettung	**		
Schilderung nonverbaler Aktionen	***	***	
Wiedergabe von Gesprächen		*	
Schilderung von Komplikationen	*	*	
Ausgefallene Einzelheiten		*	
Nebensächliche Einzelheiten	*	*	
Schilderung eigener psychischer oder physischer Vorgänge	***	**	
Schilderung psychischer oder phys. Vorgänge beim Täter oder Beteiligten			
Spontane Verbesserungen der eigenen Aussage	**		
Zugeben von Erinnerungslücken			
Einwände gegen die Richtigkeit der eigenen Aussage	*		
Selbstbelastungen			
Inschutznahme des Täters od. Beteil.			
Deliktspezifische Aussageelemente	*		

Tabelle 4-70: Ergebnisse aus den post-hoc-Verfahren (RKZ - einzeln - GLH) - einseitig

Variable (Aussagemerkmal)	wT1 – eT1	wT2 – eT2	wT1 – wT2	eT1 – eT2
Logische Konsistenz	***	***		
Widerspruchsfreie, ungeordnete Darstellung	***	***		
Quantitativer Detailreichtum	**	***		
Kontextuelle Einbettung	**	**		
Schilderung nonverbaler Aktionen	***	**	**	**
Wiedergabe von Gesprächen		*		*
Schilderung von Komplikationen		**		*
Ausgefallene Einzelheiten		**		**
Nebensächliche Einzelheiten	*	*	*	
Schilderung eigener psychischer oder physischer Vorgänge	***	***	*	***
Schilderung psychischer od. phys. Vorgänge b Täter oder Beteiligten				
Spontane Verbesserungen der eigenen Aussage	**		*	
Zugeben von Erinnerungslücken				
Einwände gegen die Richtigkeit der eigenen Aussage	*	*		
Selbstbelastungen			*	
Inschutznahme des Täters oder Beteiligten				
Deliktspezif. Aussageelemente	*	*		

wobei: ' ' = kein signifikantes Ergebnis ($p > 0,05$)' * ' = signifikantes Ergebnis ($p \leq 0,05$)' ** ' = hochsignifikantes Ergebnis ($p \leq 0,01$)' *** ' = höchstsignifikantes Ergebnis ($p \leq 0,001$)

Tabelle 4-71: Effektstärken (2) – GLH (Aussagemerkmale – einzeln)

Variable (Aussagemerkmal)	Haupteffekt ‘WAHRHEITS- STATUS’	Haupteffekt ‘BEFRAGUNGS- ZEITPUNKT’	Interaktion ‘STATUS’ - ‘ZEITPUNKT’
Logische Konsistenz	0,700	0,098	0,027
Widerspruchsfreie, ungeordnete Darstellung	0,610	0,012	0,048
Quantitativer Detailreichtum	0,408	0,118	0,001
Kontextuelle Einbettung	0,348	0,048	0,030
Schilderung nonverbaler Aktionen	0,485	0,512	0,003
Wiedergabe von Gesprächen	0,117	0,170	0,010
Schilderung von Komplikationen	0,190	0,194	0,017
Ausgefallene Einzelheiten	0,058	0,180	0,107
Nebensächliche Einzelheiten	0,223	0,227	0,018
Schilderung eigener psychischer oder physischer Vorgänge	0,533	0,321	0,001
Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	0,053	0,100	0,032
Spontane Verbesserungen der eigenen Aussage	0,260	0,122	0,064
Zugeben von Erinnerungslücken	0,033	0,001	0,064
Einwände gegen die Richtigkeit der eigenen Aussage	0,231	0,019	0,001
Selbstbelastungen	0,006	0,093	0,141
Inschutznahme des Täters oder Beteiligten	0,000	0,000	0,010
Deliktspezifische Aussageelemente	0,150	0,111	0,040

4.2.4.2 Ergebnisse zu den Merkmalskategorien und für alle Aussagemerkmale zusammen

Tabelle 4-72: Ergebnisse des ALM (3 - GLH)

Variable (Merkmalskategorie)	Haupteffekt ‘WAHRHEITS-STATUS’	Haupteffekt ‘BEFRAGUNGS-ZEITPUNKT’	Interaktion ‘STATUS’ - ‘ZEITPUNKT’
Allgemeine Merkmale	***		
Spezielle Inhalte	***	**	
Inhaltliche Besonderheiten	***	***	
Motivationsbezogene Inhalte	*		
Deliktspezifische Inhalte	*		
RKZ total	***	***	

Tabelle 4-73: Ergebnisse aus den post-hoc-Verfahren (Kategorien der Aussagemerkmale und Aussagemerkmale ‘total’ - GLH) - einseitig

Variable (Merkmalskategorie)	wT1 – eT1	wT2 – eT2	wT1 – wT2	eT1 – eT2
Allgemeine Merkmale	***	***		
Spezielle Inhalte	**	***	*	**
Inhaltliche Besonderheiten	**	***	**	***
Motivationsbezogene Inhalte	**			
Deliktspezifische Inhalte	*	*		
RKZ total	***	***	**	*

wobei: ‘ ’ = kein signifikantes Ergebnis ($p > 0,05$)

‘ * ’ = signifikantes Ergebnis ($p \leq 0,05$)

‘ ** ’ = hochsignifikantes Ergebnis ($p \leq 0,01$)

‘ *** ’ = höchstsignifikantes Ergebnis ($p \leq 0,001$)

Tabelle 4-74: Effektstärken (3) – GLH (Kategorien der Aussagemerkmale und Aussagemerkmale 'total')

Variable (Merkmalskategorie)	Haupteffekt ‘WAHRHEITS- STATUS’	Haupteffekt ‘BEFRAGUNGS- ZEITPUNKT’	Interaktion ‘STATUS’ - ‘ZEITPUNKT’
Allgemeine Merkmale	0,662	0,111	0,027
Spezielle Inhalte	0,400	0,347	0,000
Inhaltliche Besonderheiten	0,398	0,429	0,001
Motivationsbezogene Inhalte	0,229	0,086	0,076
Deliktspezifische Inhalte	0,150	0,111	0,040
RKZ total	0,546	0,366	0,023

4.2.5 Ergebnisse zur Glaubhaftigkeitseinschätzung

Im Folgenden werden die Ergebnisse der Glaubhaftigkeitseinschätzung durch den Rater (= Verfasser der vorliegenden Arbeit) zunächst über beide Befragungszeitpunkte hinweg, im Anschluss daran getrennt nach T1 und T2 dargestellt. Im Rahmen der Auswertung wurde abschließend für jede der beiden von allen Versuchspersonen erzählten Geschichten eine Einschätzung ihrer Glaubhaftigkeit auf einer 9-stufigen Skala (1-9) vorgenommen. Dabei wurden die Stufen 1 bis 4 als 'unglaubhaft', die Stufe 5 als 'unentschieden' sowie die Stufen 6 bis 9 als 'glaubhaft' klassifiziert.

Die Trefferquoten werden in allen Darstellungen in absoluten und relativen Häufigkeiten angegeben sowie im Sinne und nach Terminologie der Signalentdeckungstheorie (SDT) klassifiziert ('hits', 'correct rejections', 'misses' und 'false alarms').

4.2.5.1 Glaubhaftigkeitseinschätzung zu beiden Befragungszeitpunkten

Tabelle 4-75: Glaubhaftigkeitseinschätzung über beide Befragungszeitpunkte hinweg in absoluten und relativen Häufigkeiten

Rating (GLH-Einschätzung)	Wahrheitsstatus der Geschichte		gesamt
	erlebnis-fundiert	nicht erlebnis-fundiert	
glaubhaft	45 (45 %) <i>hits</i>	5 (5 %) <i>false alarms</i>	50 (50%)
unentschieden	0 (0 %)	1 (1 %)	1 (1 %)
nicht glaubhaft	5 (5 %) <i>misses</i>	44 (44 %) <i>correct rejections</i>	49 (49 %)
gesamt	50 (50 %)	50 (50 %)	100 (100%)

Aus oben stehender Tabelle ist ersichtlich, dass über beide Befragungszeitpunkte hinweg 89 der 100 Aussagen korrekt klassifiziert wurden, was in diesem Fall auch einer positiven Treffer-Quote von insgesamt 89 % entspricht. Dabei entfallen 45 % auf erlebnis-fundierte Aussagen, die als solche erkannt und als glaubhaft eingestuft wurden ('hits') und 44 % auf nicht erlebnis-fundierte Aussagen, die ebenfalls richtig als solche identifiziert wurden und zu recht dementsprechend für nicht glaubhaft erklärt wurden ('correct rejections'). Rein auf die wahren Geschichten bezogen entspräche dies einer positiven Trefferquote ('hits') von 90 %, bezogen auf die erfundenen Geschichten von 88 % ('correct rejections').

Der Rater war sich insgesamt über T1 und T2 hinweg nur bei einer (erfundenen) Aussage so unsicher, dass er das Urteil 'unentschieden' abgab.

Insgesamt wurden 10 der 100 Aussagen (10 %) nicht korrekt identifiziert, wobei sich die Quote der 'false alarms', also derjenigen Aussagen, die zu Unrecht als glaubhaft eingestuft wurden und die der 'misses', also wahrer Geschichten, die nicht als solche erkannt wurden, gleich darstellt. Jeweils 5 Aussagen bzw. 5 % der Aussagen wurden falsch klassifiziert. Getrennt nach erlebnis- und nicht erlebnis-fundierten Geschichten entspräche dies jeweils einer negativen Trefferquote von 10 %.

4.2.5.2 Glaubhaftigkeitseinschätzung zum ersten Befragungszeitpunkt T1

Tabelle 4-76: Glaubhaftigkeitseinschätzung zum ersten Befragungszeitpunkt T1 in absoluten und relativen Häufigkeiten

Rating (GLH-Einschätzung)	Wahrheitsstatus der Geschichte		gesamt
	erlebnis-fundiert	nicht erlebnis-fundiert	
glaubhaft	22 (44 %) <i>hits</i>	3 (6 %) <i>false alarms</i>	25 (50%)
unentschieden	0 (0 %)	0 (0 %)	0 (0 %)
nicht glaubhaft	3 (6 %) <i>misses</i>	22 (44 %) <i>correct rejections</i>	25 (50 %)
gesamt	25 (50 %)	25 (50 %)	50 (100%)

Zum ersten Befragungszeitpunkt T1 wurden insgesamt 44 von 50 Aussagen (88 %) korrekt identifiziert und klassifiziert werden, wobei 22 der 50 Aussagen (44 %) als erlebnis-fundiert erkannt und für glaubhaft befunden wurden ('hits'), sowie ebenfalls 22 der 50 als nicht erlebnis-fundiert und für unglaublich ('correct rejections'). Wiederum nur auf die wahren bzw. die erfundenen Geschichten bezogen, würde dies einer positiven Trefferquote ('hits' bzw. 'correct rejections') von jeweils 88 % entsprechen.

Zu diesem Befragungszeitpunkt konnten insgesamt 6 von 50 Aussagen nicht richtig zugeordnet bzw. deren tatsächlicher Wahrheitsstatus ermittelt werden, was einer relativen Quote von 12 % entspricht. Davon entfielen wiederum jeweils die Hälfte, also 3 Geschichten bzw. 6 % auf die 'misses' und 'false alarms', wurden also entweder trotz Erlebnishintergrund als nicht glaubhaft eingestuft oder ohne Erlebnishintergrund für glaubhaft gehalten. Analog zu obigen Ausführungen erhielt man – getrennt für wahre bzw. erfundene Geschichten - negative Trefferquoten von jeweils 12 %.

4.2.5.3 Glaubhaftigkeitseinschätzung zum zweiten Befragungszeitpunkt T2

Tabelle 4-77: Glaubhaftigkeitseinschätzung zum zweiten Befragungszeitpunkt T2 in absoluten und relativen Häufigkeiten

Rating (GLH-Einschätzung)	Wahrheitsstatus der Geschichte		gesamt
	erlebnis-fundiert	nicht erlebnis-fundiert	
glaubhaft	23 (46 %) <i>hits</i>	2 (4 %) <i>false alarms</i>	25 (50%)
unentschieden	0 (0 %)	1 (2 %)	1 (2 %)
nicht glaubhaft	2 (4 %) <i>misses</i>	22 (44 %) <i>correct rejections</i>	24 (48 %)
gesamt	25 (50 %)	25 (50 %)	50 (100%)

Über alle Aussagen betrachtet verändert sich das Ergebnis zu T2 nur wenig gegenüber der Einschätzung von T1.

Es werden insgesamt 45 der 50 Aussagen (90 %) richtig zugeordnet. Im Vergleich zum ersten Befragungszeitpunkt T1 steigt die 'hit - rate' auf 46 %, während die Rate der 'correct rejections' gleich bei 44 % bleibt. Betrachtet man nur die erlebnis-basierten Aussagen ergäbe sich zu diesem Befragungszeitpunkt eine positive Trefferquote von 92 % ('hits') bzw. für die nicht erlebnis-basierten Geschichten von wiederum 88 % ('correct rejections').

Lediglich bei einer Geschichte (2 %) konnte sich der Rater nicht festlegen.

Die Gesamtzahl und -rate der nicht richtig identifizierten Aussagen sinkt im Vergleich zu T1 um 2 auf jetzt 4 bzw. von 12 % auf 8 %. Dabei ist das Verhältnis von 'misses' und 'false alarms' wiederum ausgeglichen bei 2 bzw. 4 % nicht korrekt klassifizierter Geschichten. Getrennt nach wahren und erfundenen Geschichten betragen die 'misses' und 'false alarms' (negative Trefferquoten) jeweils 8 %.

Insgesamt kann festgehalten werden, dass die Geschichten bzw. deren jeweilige Erlebnisfundierung über beide Befragungszeitpunkte hinweg, sowie auch zu beiden Befragungszeitpunkten einzeln sehr gut und in etwa im selben Ausmaß ermittelt werden und im Rating eine entsprechende Glaubhaftigkeitseinschätzung vorgenommen werden konnte.

4.2.6 Klassifikation der Glaubhaftigkeit anhand der erhobenen Aussagemerkmale mittels Diskriminanzanalyse

4.2.6.1 Analyse der Gruppenunterschiede und Prüfung der Trennfähigkeit der Indikatoren zu beiden Befragungszeitpunkten

Die Prüfung der Trennfähigkeit der Indikatoren, also der Aussagemerkmale dient der Klärung der Frage, welche der Indikatoren zur Identifikation von Wahr- bzw. Falschaussagen (Aussagen mit Erlebnishintergrund bzw. ohne) für die Unterscheidung der wahren von den erfundenen Geschichten „diskriminatorisch“ bedeutsam sind, d.h. zur Trennung des Wahrheitsstatus beitragen.

Mit Hilfe von Wilks' Lambda lässt sich für jedes Aussagemerkmal zunächst isoliert dessen Trennfähigkeit ermitteln (vgl. BACKHAUS et al., 2000). In Analogie zu einer einfachen Varianzanalyse wird für jeden Indikator der Mittelwert der Ausprägungen, den die Versuchspersonen innerhalb der wahren Geschichten auf diesem Indikator aufweisen, mit dem Mittelwert der Ausprägungen der Versuchspersonen innerhalb der erfundenen Geschichten verglichen. Indikatoren, die nicht zur Unterscheidung von erlebnisbasierten von nicht erlebnisbasierten Aussagen beitragen, unterscheiden sich in ihren Mittelwerten nicht signifikant.

Befragungszeitpunkt T1*Tabelle 4-78: Gleichheitstest der Gruppenmittelwerte zum ersten Befragungszeitpunkt T1*

AUSSAGEMERKMAL (Realkennzeichen)	Wilk's Lambda	F	df 1	df 2	p
Logische Konsistenz	0,378	79,085	1	48	< 0,001
Widerspruchsfreie, ungeordnete Darstellung	0,540	40,960	1	48	< 0,001
Quantitativer Detailreichtum	0,816	10,811	1	48	0,002
Kontextuelle Einbettung	0,789	12,834	1	48	0,001
Schilderung nonverbaler Aktionen	0,760	15,120	1	48	< 0,001
Wiedergabe von Gesprächen	0,975	1,223	1	48	0,274
Schilderung von Komplikationen im Handlungsverlauf	0,949	2,592	1	48	0,114
Ausgefallene Details	1,000	0,000	1	48	1,000
Nebensächliche Details	0,927	3,756	1	48	0,059
Phänomengemäße Schilderung unverstandener Handlungselemente	nicht berech.	nicht berech.	1	48	nicht berech.
Inhaltliche Verschachtelungen	0,993	0,343	1	48	0,561
Schilderung eigener psychischer oder physischer Vorgänge	0,681	22,507	1	48	< 0,001
Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	0,969	1,556	1	48	0,218
Spontane Verbesserungen der eigenen Aussage	0,859	7,882	1	48	0,007
Zugeben von Erinnerungslücken	0,960	2,021	1	48	0,162
Einwände gegen die Richtigkeit der eigenen Aussage	0,932	3,479	1	48	0,068
Selbstbelastungen	0,971	1,430	1	48	0,238
Inschutznahme des Täters oder Beteiligten	0,893	5,731	1	48	0,021
Deliktspezifische Aussageelemente	0,929	3,675	1	48	0,061

Bei Durchführung einer einfaktoriellen Varianzanalyse (ANOVA) zum Test der Nullhypothese der Gleichheit der Gruppenmittelwerte der Aussagemerkmale ergab sich für den ersten Befragungszeitpunkt T1 für folgende Kriterien bei einer Irrtumswahrscheinlichkeit von 5 % kein signifikanter Mittelwertsunterschied (vgl. *Tabelle 4-78*):

‘Wiedergabe von Gesprächen’, ‘Schilderung von Komplikationen im Handlungsverlauf’, ‘Ausgefallene Details’, ‘Nebensächliche Details’, ‘Phänomengemäße Schilderung unverstandener Handlungselemente’, ‘Inhaltliche Verschachtelungen’, ‘Spontane Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten’, ‘Zugeben von Erinnerungslücken’, Einwände gegen die Richtigkeit der eigenen Aussage’, ‘Selbstbelastungen’ sowie ‘Deliktspezifische Aussageelemente’.

Diese Indikatoren tragen zum Zeitpunkt T1 folglich zum einen nicht zur Trennung von erlebnisfundierte und nicht erlebnisfundierte Aussagen bei, zum anderen auch nicht zu einer erhöhten Varianzaufklärung durch das Diskriminanzkriterium. Daher können sie im Folgenden für die Bildung der Diskriminanzfunktionen vernachlässigt werden. Alle anderen Real-kennzeichen trennen jedoch hier signifikant und können somit als Grundlage für die weitere Analyse herangezogen werden.

Befragungszeitpunkt T2*Tabelle 4-79: Gleichheitstest der Gruppenmittelwerte zum ersten Befragungszeitpunkt T2*

AUSSAGEMERKMAL (Realkennzeichen)	Wilk's Lambda	F	df 1	df 2	p
Logische Konsistenz	0,475	53,159	1	48	< 0,001
Widerspruchsfreie, ungeordnete Darstellung	0,698	21,676	1	48	< 0,001
Quantitativer Detailreichtum	0,820	10,566	1	48	0,002
Kontextuelle Einbettung	0,886	6,193	1	48	0,016
Schilderung nonverbaler Aktionen	0,830	9,823	1	48	0,003
Wiedergabe von Gesprächen	0,950	2,519	1	48	0,119
Schilderung von Komplikationen im Handlungsverlauf	0,872	7,053	1	48	0,011
Ausgefallene Details	0,860	7,837	1	48	0,007
Nebensächliche Details	0,949	2,560	1	48	0,116
Phänomengemäße Schilderung unverstandener Handlungselemente	0,980	1,000	1	48	0,322
Inhaltliche Verschachtelungen	0,993	0,343	1	48	0,561
Schilderung eigener psychischer oder physischer Vorgänge	0,726	18,112	1	48	< 0,001
Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	0,992	0,388	1	48	0,536
Spontane Verbesserungen der eigenen Aussage	0,949	2,592	1	48	0,114
Zugeben von Erinnerungslücken	0,999	0,036	1	48	0,850
Einwände gegen die Richtigkeit der eigenen Aussage	0,902	5,189	1	48	0,027
Selbstbelastungen	0,988	0,603	1	48	0,441
Inschutznahme des Täters oder Beteiligten	0,998	0,091	1	48	0,764
Deliktspezifische Aussageelemente	0,923	4,000	1	48	0,051

Zu diesem zweiten Befragungszeitpunkt T2 ergab sich nach Durchführung einer einfaktoriellen Varianzanalyse (ANOVA) zum Test der Nullhypothese der Gleichheit der Gruppenmittelwerte der Realkennzeichen für folgende Kriterien bei einer Irrtumswahrscheinlichkeit von 5 % kein signifikanter Mittelwertsunterschied (vgl. *Tabelle 4-79*):

‘Wiedergabe von Gesprächen’, ‘Nebensächliche Details’, ‘Phänomengemäße Schilderung unverstandener Handlungselemente’, ‘Inhaltliche Verschachtelungen’, ‘Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten’, ‘Spontane Verbesserungen der eigenen Aussage’, ‘Zugeben von Erinnerungslücken’, ‘Selbstbelastungen’, ‘Inschutznahme des Täters oder Beteiligten’ sowie ‘Deliktspezifische Aussageelemente’.

Diese Indikatoren tragen zum Zeitpunkt T2 folglich nicht zur Trennung der Gruppen ‘wahr’ versus ‘erfunden’ bei. Da diese nicht zu einer erhöhten Varianzaufklärung durch das Diskriminanzkriterium und somit nicht zu einer besseren Trennung zwischen den Gruppen beitragen, können sie im Folgenden für die Bildung der Diskriminanzfunktionen vernachlässigt werden. Alle anderen Realkennzeichen trennen jedoch hier signifikant und können somit als Grundlage für die weitere Analyse herangezogen werden.

Die Diskriminanzanalyse liefert für die identifizierten Realkennzeichen bzw. deren Indikatoren darüber hinaus die geschätzten Diskriminanz- bzw. Gewichtungskoeffizienten. Die identifizierten Glaubwürdigkeitskriterien und deren zugehörige *unstandardisierte* Gewichtungskoeffizienten bilden zusammen die jeweiligen Diskriminanzfunktionen für die jeweiligen Befragungszeitpunkte.

Für die Beurteilung der diskriminatorischen Bedeutung der Indikatoren benötigt man nach BACKHAUS et al. (2000) jedoch die *standardisierten* Gewichtungskoeffizienten. Die Größe eines Diskriminanzkoeffizienten ist unter anderem von der Standardabweichung des zugehörigen Indikators abhängig und damit von eventuell willkürlichen Skalierungseffekten. Um diesen Effekt auszuschalten, werden die Diskriminanzkoeffizienten mit den Standardabweichungen des jeweiligen Indikators multipliziert (vgl. BACKHAUS et al., 2000). Die standardisierten Gewichtungskoeffizienten geben Auskunft über die „Wichtigkeit“ der verschiedenen Indikatoren für die Trennung der Gruppen (vgl. BACKHAUS et al., 2000). Sie stellen also gewissermaßen Faktoren dar, mit denen die jeweiligen Mittelwerte der Realkennzeichen gewichtet werden. Je höher der Betrag des jeweiligen Gewichtungskoeffizienten ist (unabhängig vom Vorzeichen), umso wichtiger ist dieser Indikator für die Trennung der Gruppen.

Standardisierte kanonische Diskriminanzfunktionskoeffizienten

Tabelle 4-80: Standardisierte kanonische Diskriminanzfunktionskoeffizienten

AUSSAGEMERKMAL (Realkennzeichen)	Befragungszeitpunkt	
	T1	T2
Logische Konsistenz	0,779	0,921
Widerspruchsfreie, ungeordnete Darstellung	0,418	0,508
Quantitativer Detailreichtum	-0,396	-0,281
Kontextuelle Einbettung	0,045	-0,566
Schilderung nonverbaler Aktionen	0,289	0,408
Wiedergabe von Gesprächen	-	-
Schilderung von Komplikationen im Handlungsverlauf	-	-0,471
Ausgefallene Details	-	0,204
Nebensächliche Details	-	-
Phänomengemäße Schilderung unverstandener Handlungselemente	-	-
Inhaltliche Verschachtelungen	-	-
Schilderung eigener psychischer oder physischer Vorgänge	0,218	0,564
Schilderung psychischer oder physischer Vorgänge beim Täter / Beteiligten	-	-
Spontane Verbesserungen der eigenen Aussage	0,118	-
Zugeben von Erinnerungslücken	-	-
Einwände gegen die Richtigkeit der eigenen Aussage	-	0,346
Selbstbelastungen	-	-
Inschutznahme des Täters oder Beteiligten	-0,002	-
Deliktspezifische Aussageelemente	-	-

Tabelle 4-80 zeigt zunächst ungeordnet die jeweiligen standardisierten kanonischen Diskriminanzfunktionskoeffizienten. Bringt man diese dann in eine Rangreihe - d.h. dasjenige Realkennzeichen mit dem größten Gewicht, also der größten Aussagekraft steht an erster Stelle – so ergibt sich getrennt nach Befragungszeitpunkten nachstehende Tabelle der diskriminatorisch bedeutsamen Aussagemerkmale.

4.2.6.2 Die diskriminatorisch bedeutsamen Aussagemerkmale zu T1 und T2

Tabelle 4-81: Rangfolge der gewichteten diskriminatorisch bedeutsamen Aussagemerkmale zu beiden Befragungszeitpunkten T1 und T2

Befragungszeitpunkt T1		Befragungszeitpunkt T2	
Rangplatz	Aussagemerkmal	Rangplatz	Aussagemerkmal
1	Logische Konsistenz	1	Logische Konsistenz
2	Widerspruchsfreie, ungeordnete Darstellung	2	Kontextuelle Einbettung
3	Quantitativer Detailreichtum	3	Schilderung eigener psychischer oder physischer Vorgänge
4	Schilderung nonverbaler Aktionen	4	Widerspruchsfreie, ungeordnete Darstellung
5	Schilderung eigener psychischer oder physischer Vorgänge	5	Schilderung von Komplikationen im Handlungsverlauf
6	Spontane Verbesserungen der eigenen Aussage	6	Schilderung nonverbaler Aktionen
7	Kontextuelle Einbettung	7	Einwände gegen die Richtigkeit der eigenen Aussage
8	Inschutznahme des Täters oder Beteiligten	8	Quantitativer Detailreichtum
		9	Ausgefallene Details

Zusammenfassend lässt sich somit die Schlussfolgerung ziehen, dass die oben stehenden, verbleibenden acht Aussagemerkmale zum ersten Befragungszeitpunkt T1 und neun Aussagemerkmale zum zweiten Befragungszeitpunkt T2 alleine hinreichend für eine signifikante Trennung zwischen wahren (erlebnisbasierten) und erfundenen (ohne Erlebnishintergrund) Aussagen sind.

4.2.6.3 Die Prüfung der Diskriminanzfunktionen

Neben der Prüfung der isolierten Trennfähigkeit der Indikatoren wird ebenfalls überprüft, ob sich die beiden Diskriminanzgruppen (erlebnisfundierte und nicht erlebnisfundierte Aussagen) über alle identifizierten Aussagemerkmale hinweg signifikant unterscheiden. Dies erfolgt über die statistische Signifikanzprüfung der Diskriminanzfunktionen.

Zur Prüfung der Diskriminanzfunktionen wird auf den sog. **Eigenwert Υ** als Maximalwert des Diskriminanzkriteriums zurückgegriffen, der ein Maß für die Unterschiedlichkeit der Gruppen, wie auch für die Güte (Trennkraft) einer Diskriminanzfunktion darstellt. Zur Normierung des Eigenwerts auf Werte zwischen 0 und 1 wird folgende Transformation vorgenommen:

$$L = 1 / (1 + \Upsilon)$$

Dieser Ausdruck wird als **Wilks' Lambda** bezeichnet. Wilk's Lambda ist ein so genanntes „inverses Gütemaß“, d.h. kleinere Werte bedeuten eine höhere Unterschiedlichkeit der Gruppen bzw. eine höhere Trennkraft der Diskriminanzfunktionen und umgekehrt. Die Bedeutung von Wilks' Lambda liegt nach BACKHAUS et al. (2000) darin, dass es sich in eine probabilistische Variable transformieren lässt und somit Wahrscheinlichkeitsaussagen über die Unterschiedlichkeit der Gruppen erlaubt. Dies ermöglicht eine statistische Signifikanzprüfung der Diskriminanzfunktionen über einen χ^2 - Test. Die Signifikanzprüfung beinhaltet die Überprüfung der Nullhypothese, dass die beiden Gruppen sich nicht signifikant unterscheiden. Je größer der χ^2 - Wert ist, desto geringer ist die Wahrscheinlichkeit, dass die Nullhypothese zutrifft (vgl. BACKHAUS et al., 2000).

Tabelle 4-82: Statistische Signifikanzprüfung der Diskriminanzfunktionen für beide Befragungszeitpunkte T1 und T2

SIGNIFIKANZPRÜFUNG DER DISKRIMINANZFUNKTION: T1			
Wilk's Lambda	CHI-Quadrat	df	Signifikanz
0,284	49,077	18	< 0,001
SIGNIFIKANZPRÜFUNG DER DISKRIMINANZFUNKTION: T2			
Wilk's Lamba	Chi-Quadrat	df	Signifikanz
0,251	53,255	19	< 0,001

Tabelle 4-82 verdeutlicht, dass die Nullhypothese zu beiden Befragungszeitpunkten mit einer Irrtumswahrscheinlichkeit, die kleiner ist als 0,1 % zu verwerfen ist. Zu beiden Zeitpunkten ergeben sich höchstsignifikante Ergebnisse ($p < 0,001$).

Die ermittelten Diskriminanzfunktionen trennen somit zu beiden Zeitpunkten sehr deutlich zwischen den Versuchsgruppen; die Aussagen mit Erlebnishintergrund lassen sich also mittels der beiden identifizierten Diskriminanzfunktionen zu beiden Befragungszeitpunkten eindeutig von den Aussagen ohne Erlebnishintergrund unterscheiden.

4.2.6.4 Der kritische Diskriminanzwert

Mit Hilfe des kritischen Diskriminanzwertes kann eine diskriminanzanalytische Klassifikation der Versuchspersonen vorgenommen werden. Hierfür werden innerhalb der Gruppe der wahren und der erfundenen Geschichten neben den individuellen Diskriminanzwerten zunächst die Mittelwerte der Diskriminanzwerte der einzelnen Versuchspersonen bestimmt. Der kritische Diskriminanzwert berechnet sich (vereinfacht) über folgende Formel:

$$Y^* = (\bar{Y}_W + \bar{Y}_E) / 2$$

Y^* = kritischer Diskriminanzwert

Y_W = Mittelwert der Diskriminanzwerte
innerhalb der wahren Geschichten
(erlebnisfundiert)

Y_E = Mittelwert der Diskriminanzwerte
innerhalb der erfundenen Geschichte
(nicht erlebnisfundiert)

Tabelle 4-83 zeigt die Gruppenmittelwerte der individuellen Diskriminanzwerte für die beiden Versuchsgruppen sowie jeweils den kritischen Diskriminanzwert Y^* .

Tabelle 4-83: Gruppenmittelwerte und kritische Diskriminanzwerte für die Diskriminanzgruppen zu beiden Befragungszeitpunkten T1 und T2

Gruppenmittel	T1	T2
erlebnisfundierte Geschichten	1,298	0,847
nicht erlebnisfundierte	-1,165	-0,724
Kritischer Diskriminanzwert	0,004	0,026

Liegt der Diskriminanzwert einer Versuchsperson unter dem kritischen Wert Y^* , wird sie durch die Diskriminanzfunktion der Gruppe der erlebnisfundierten Aussagen zugeordnet; liegt er darüber, der Gruppe der nicht erlebnisfundierten. D.h. mittels der ermittelten Diskriminanzfunktionen lässt sich in Verbindung mit dem jeweiligen kritischen Diskriminanzwert (zu Befragungszeitpunkt T1 und T2) eine Zugehörigkeit zu einer der beiden Versuchsgruppen vornehmen.

4.2.6.5 Klassifizierungsergebnisse

Neben dem Diskriminanzkriterium gibt es auch eine zweite Möglichkeit zur Prüfung der Diskriminanzfunktion. Man vergleicht die durch die Diskriminanzfunktion bewirkte Klassifizierung der Untersuchungsobjekte mit deren tatsächlicher Gruppenzugehörigkeit (vgl. BACKHAUS et al., 2000, S. 170). Die Häufigkeiten der korrekt und falsch klassifizierten Elemente für die verschiedenen Gruppen lassen sich übersichtlich in einer so genannten Klassifikationsmatrix zusammenfassen.

Analog zu den Kreuztabellen im Kapitel 'Einschätzung der Glaubhaftigkeit' (4.2.5) und der SDT-Terminologie finden sich in den Hauptdiagonalen der *Tabellen 4-84* und *4-85* die absoluten Fallzahlen bzw. die relativen Häufigkeiten (in Klammern) der korrekt klassifizierten Elemente jeder Gruppe; in den übrigen Feldern die der falsch klassifizierten.

Für den ersten Befragungszeitpunkt T1 ergibt sich schließlich folgendes Ergebnis:

Tabelle 4-84: Klassifikationsmatrix (Klassifizierungsergebnisse) zu T1

Tatsächlicher Wahrheitsstatus	Prognostizierter Wahrheitsstatus		gesamt
	erlebnisfundiert	nicht erlebnisfundiert	
erlebnisfundiert	24 (96 %)	1 (4 %)	25 (100 %)
nicht erlebnisfundiert	1 (4 %)	24 (96 %)	25 (100 %)

Aus der obigen *Tabelle 4-84* ist ersichtlich, dass bezüglich des ersten Befragungszeitpunktes T1 durch die Diskriminanzfunktion jeweils 96 % der erlebnisbasierten sowie der nicht erlebnisbasierten Aussagen korrekt klassifiziert werden konnten.

Insgesamt wurden letztendlich entsprechend 96 % der ursprünglich gruppierten Fälle korrekt klassifiziert, dementsprechend 4 % nicht korrekt. Dies ergibt sich aus dem Mittel der jeweiligen Diagonalen.

Für den zweiten Befragungszeitpunkt T2 stellen sich die Klassifizierungsergebnisse der Diskriminanzfunktion wie folgt dar:

Tabelle 4-85: Klassifikationsmatrix (Klassifizierungsergebnisse) zu T2

Tatsächlicher Wahrheitsstatus	Prognostizierter Wahrheitsstatus		gesamt
	erlebnisfundiert	nicht erlebnisfundiert	
erlebnisfundiert	23 (92 %)	2 (8 %)	25 (100 %)
nicht erlebnisfundiert	0 (0 %)	25 (100 %)	25 (100 %)

Tabelle 4-85 verdeutlicht, dass sich zum zweiten Zeitpunkt T2 92 % `Hits` ergeben, d.h. von den erlebnisfundierten Aussagen konnten 23 von 25 Aussagen korrekt als wahr klassifiziert werden. In der Gruppe der nicht erlebnisfundierten konnten alle 25 erfundenen Geschichten (100 %) korrekt als solche klassifiziert werden (`Correct Rejections`).

Insgesamt konnten für den zweiten Befragungszeitpunkt T2 durch die Diskriminanzfunktion 96 % der ursprünglich gruppierten Fälle korrekt klassifiziert werden. 4 % wurden falsch klassifiziert, was sich wiederum aus den mittleren Werten der jeweiligen Diagonalen ergibt.

4.3 Zusammenfassung der Ergebnisse

4.3.1 Erinnerungsleistung und source- bzw. reality monitoring

Es werden im Folgenden jeweils zunächst die Ergebnisse über beide Dia- bzw. Unfall-Serien hinweg zusammenfassend dargestellt, danach jeweils zuerst für die Serie 'PKW-PKW' und dann für die Serie 'PKW-Fahrrad'.

4.3.1.1 Ergebnisse der Unterschiede zwischen den Versuchsgruppen

Die folgenden zusammenfassenden Befunde [a-a-...) bis a-c-...)] beziehen sich auf den ersten Befragungszeitpunkt T1.

- a-a-a) Zum ersten Befragungszeitpunkt T1 berichtet die 'reality-monitoring' - Gruppe zwar erkennbar mehr, jedoch nicht signifikant mehr korrekte Details insgesamt als die externe 'source-monitoring'- Gruppe, auf Nachfrage sind es allerdings signifikant mehr korrekte Details. Im freien Bericht sind die diesbezüglichen Werte zu T1 nahezu identisch.
 - a-a-b) Die Gruppe 'external-internal' begeht zum ersten Befragungszeitpunkt insgesamt signifikant mehr Fehler als die Gruppe 'external-external'.
 - a-a-c) Hinsichtlich der Quellenverwechslungen insgesamt, innerhalb und zwischen den Serien sind die Unterschiede sowie die einzelnen Werte recht gering. Lediglich zwischen den Serien finden sich in der Versuchsbedingung 'external-external' signifikant mehr Quellenverwechslungen als in der Bedingung 'external-internal'.
 - a-a-d) Die Versuchspersonen der Bedingung 'external-external' begehen ebenso signifikant mehr schematypische Intrusionen als die Jugendlichen der Bedingung 'external-internal'.
 - a-a-e) Konfabulationen treten zu diesem ersten Befragungszeitpunkt jedoch sogar hochsignifikant mehr in der 'reality-monitoring' - Gruppe auf.
 - a-a-f) Was allgemeine Erinnerungsfehler betrifft, so erweisen sich die beobachteten Unterschiede zwischen den Gruppen zu T1 als statistisch unbedeutend.
- a-b-a) Bei einzelner Betrachtung der Dia-Serie 'PKW-PKW' ist festzuhalten, dass die RM – Gruppe insgesamt sowie auf Nachfrage deutlich mehr korrekte Details erinnert und berichtet als die ESM – Gruppe. Im freien Bericht sind dies nur tendenziell mehr.

- a-b-b) Hinsichtlich der insgesamt begangenen Fehler, der allgemeinen Erinnerungsfehler, der berichteten schematischen Intrusionen sowie der Konfabulationen ließen sich für die Dia-Serie 'PKW-PKW' zwischen den Versuchsgruppen im Ergebnis keine signifikanten Unterschiede feststellen.
- a-b-c) Wie auch über beide Serien hinweg betrachtet ergibt sich auch hier lediglich zwischen den Serien ein bedeutsamer Unterschied zwischen den Jugendlichen der Versuchsgruppen hinsichtlich aufgetretener Quellenverwechslungen, wiederum dergestalt, dass solche Fehler in der ESM – Gruppe deutlich häufiger zu finden sind als in der RM – Gruppe. Dort fanden sich überhaupt keine Quellenverwechslungen zwischen den Serien. Das Ausmaß der Quellenverwechslungen insgesamt sowie innerhalb der Dia-Serie war in beiden Versuchsgruppen erneut statistisch unbedeutend.
- a-c-a) Völlig gegensätzlich verhalten sich die Beobachtungen und Ergebnisse für die Dia- bzw. Unfall-Serie 'PKW-Fahrrad'.
- Hier fanden sich weder hinsichtlich korrekt berichteter Details noch irgendeiner Art von Quellenverwechslungen nennenswerte Unterschiede oder signifikante Ergebnisse zwischen den Versuchsbedingungen 'external-external' und 'external-internal'.
- a-c-b) Allerdings scheinen die Gedächtnisleistungen der Versuchspersonen der Gruppe RM hinsichtlich der übrigen Fehlerarten (allgemeine Erinnerungsfehler und Konfabulationen) sowie auch der Gesamtfehlerzahl weitaus störanfälliger zu sein als die der Gruppe ESM. Es zeigen sich signifikante und hochsignifikante Unterschiede.
- a-c-c) Im Ergebnis ebenfalls signifikant erweist sich der Befund der schematischen Intrusionen, jedoch kommen diese hier in der RM – Gruppe weniger oft vor als in der ESM – Gruppe.

Die folgenden zusammenfassenden Befunde [b-a-...) bis b-c-...)] beziehen sich auf einen zweiten Befragungszeitpunkt, wenn schon eine erste Befragung voranging.

- b-a-a) Wie zu T1 lassen sich auch hier hinsichtlich der insgesamt und auf Nachfrage korrekt berichteten Details keine signifikanten Unterschiede zwischen den beiden Versuchsgruppen feststellen. Dies gilt hier außerdem für die korrekten Details aus dem freien Bericht. Die Unterschiede zeigen sich hier geringer als zu T1.
- b-a-b) Auch was die Gesamtfehlerzahl angeht, zeigt sich ein analoges Ergebnis zum ersten Befragungszeitpunkt. Den Jugendlichen der Versuchsgruppe 'external-internal' unterlaufen hier insgesamt sehr viel mehr Fehler als den Jugendlichen der Versuchsgruppe 'external-external'.

- b-a-c) Wiederum finden sich keinerlei signifikante Veränderungen bezüglich der Quellenverwechslungen insgesamt, innerhalb oder zwischen den Serien.
- b-a-d) Hinsichtlich des Vorkommens von schematypischen Intrusionen sowie auch von Konfabulationen zeigen sich zu diesem Befragungszeitpunkt jeweils analoge Befunde wie zum ersten Befragungszeitpunkt:
Intrusionen kommen hier deutlich mehr in der externalen 'source-monitoring' - Gruppe vor als in der 'reality-monitoring' - Gruppe. Umgekehrt konfabulieren die Jugendlichen der Versuchsbedingung 'external-internal' deutlich mehr als die Jugendlichen der Vergleichsgruppe.
- b-a-e) Allgemeine Erinnerungsfehler scheinen hier zu einem zweiten Befragungszeitpunkt im Gegensatz zum ersten an Bedeutung zu gewinnen. Die zu T1 zwar schon vorhandene Tendenz zeigt sich hier in einem auch statistisch höchst bedeutsamen höheren Fehlerwert der RM – Gruppe gegenüber der ESM - Gruppe.
- b-b-a) Für die Dia-Serie 'PKW-PKW' scheint hinsichtlich des Erinnerns korrekter Details auch nach acht Wochen dasselbe zu gelten wie nach einer Woche zu T1. Insgesamt sowie auf Nachfrage berichten die Jugendlichen der RM – Gruppe wesentlich mehr Details als die der ESM – Gruppe. Im freien Bericht sind es wiederum auch nur tendenziell mehr.
- b-b-b) Anders als zu T1 ist nun die Gesamtfehlerrate sowie die Anzahl im Bereich der allgemeinen Erinnerungsfehler in der Bedingung 'external-internal' deutlich, d.h. hochsignifikant höher als in der Bedingung 'external-external'.
- b-b-c) Ebenso bestätigt bzw. wiederholt sich das Ergebnis (bei vergleichender Betrachtung von T1 und T2) bezüglich der Unterschiede der zwischen den Serien begangenen Quellenverwechslungen. Auch nach acht Wochen sind diese in der Gruppe RM nicht zu finden, im Gegensatz zur Gruppe ESM, was sich in einem erneut signifikantem Ergebnis widerspiegelt.
- b-b-d) Anhand sonstiger Quellenverwechslungen, Intrusionen und Konfabulationen können die beiden Versuchsgruppen nicht statistisch bedeutsam unterschieden werden.
- b-c-a) Das zwar tendenziell schon von T1 bekannte Ergebnis zeigt sich nun bei erneuter Befragung nach acht Wochen auch statistisch zumindest was die insgesamt und im free recall berichteten Details angeht. Bezüglich der Dia- bzw. Unfall-Serie 'PKW-Fahrrad' erinnern und berichten die Versuchspersonen der RM – Gruppe signifikant weniger als diejenigen der ESM – Gruppe. Auf Nachfrage hin bleibt der Unterschied allerdings nur richtungsgemäß tendenziell erkennbar.

- b-c-b) Zu T2 können ebenfalls die statistischen Befunde hinsichtlich der Gesamtfehlerzahl, der Zahl der allgemeinen Erinnerungsfehler, der schematischen Intrusionen sowie der Konfabulationen auch exakt (d.h. auch bezüglich der Richtung der Mittelwertsunterschiede) nicht nur bestätigt und als wiederholt angesehen werden; die jeweiligen Unterschiede sind für die Serien 'PKW-Fahrrad' deutlich größer als zu T1 und zeigen sich in entsprechend hoch- und höchstsignifikanten Ergebnissen.
- b-c-c) Hinsichtlich der Quellenverwechslungen ergibt sich ebenfalls wie zu T1 keinerlei signifikanter Unterschied zwischen den Gruppen.

Nachstehende zusammenfassende Feststellungen [c-a-...) bis c-c-...)] beziehen sich auf einen Befragungszeitpunkt NT2, jedoch ohne dass zuvor eine erste Befragung stattgefunden hat:

- c-a-a) In fast exakt analoger Weise zu Befragungszeitpunkt T2 gestalten sich die Ergebnisse bezüglich der korrekt berichteten Details zu NT2. Auch zu diesem Befragungszeitpunkt scheinen diese hier weder insgesamt, im freien Bericht noch auf Nachfrage eine geeignete Diskriminanzfunktion aufzuweisen. Es ergeben sich keine signifikanten Unterschiede zwischen den Versuchsbedingungen.
- c-a-b) Dahingegen können die beiden Versuchsgruppen bezüglich der Gesamtfehlerzahl ein weiteres Mal sehr gut unterschieden werden. Wiederum finden sich hier in der 'reality monitoring' - Gruppe wesentlich mehr Fehler als in der externalen 'source-monitoring' - Gruppe. Der Unterschied ist zu diesem Befragungszeitpunkt im Vergleich zu den beiden anderen mit Abstand am größten.
- In der Gesamtbetrachtung lässt sich jeweils ein Anstieg der absoluten Gesamtfehlerzahl in der RM – Gruppe sowie auch des Unterschieds zwischen den Gruppen von T1 nach T2 nach NT2 konstatieren - vergleicht man die drei Befragungszeitpunkte diesbezüglich direkt miteinander.
- c-a-c) Ein weiteres Mal können die Versuchsgruppen anhand der Quellenverwechslungen (insgesamt, innerhalb und zwischen den Serien) nicht unterschieden werden.
- c-a-d) Im Gegensatz zu den Befragungszeitpunkten T1 und T2 finden sich zu NT2 deutlich und signifikant mehr schematische Intrusionen in der Versuchsbedingung 'external-internal' als in der Bedingung 'external-external'.
- c-a-e) Wiederum und tendenziell mehr konfabuliert hingegen wird – wie zu T1 und T2 auch – in der Gruppe 'external-internal' gegenüber der Gruppe 'external-external'. Allerdings erweist sich der Unterschied hier statistisch gesehen als nicht-signifikant.

- c-a-f) In Analogie zu T2 begehen die Versuchspersonen der 'reality-monitoring' - Gruppe wesentlich mehr – statistisch höchstsignifikant – allgemeine Erinnerungsfehler als diejenigen der externalen 'source-monitoring' - Gruppe.
- c-b-a) Bei einer ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen ergaben sich nun sowohl für die Gesamtzahl als auch für die im freien Bericht sowie auf Nachfrage korrekt berichteten Details sehr deutliche Unterschiede zwischen ESM- und RM – Gruppe, wobei letztere sehr viel mehr Einzelheiten erinnert bzw. äußert.
- c-b-b) Im Gegensatz zu T1 (nach einer Woche), jedoch identisch zu T2 (nach acht Wochen) begeht die Gruppe RM offensichtlich bei einer ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen wesentlich mehr Fehler insgesamt wie auch im Bereich der allgemeinen Erinnerungsfehler als die Gruppe ESM.
- c-b-c) Wie auch schon zu T1 und auch zu T2 scheinen innerhalb der Dia-Serie 'PKW-PKW' schematypische Intrusionen sowie Konfabulationen auch zu diesem Befragungszeitpunkt keine entscheidende Rolle bei der Differenzierung zwischen den beiden Versuchsbedingungen zu spielen. Die genannten Fehlerarten kommen in beiden Gruppen in etwa im selben Ausmaß vor, es konnten somit auch hier keine signifikanten Unterschiede festgestellt werden.
- c-b-d) Ebenfalls ein statistisch identischer Befund ließ sich erneut, wie auch zu T1 und T2, für den Bereich Quellenverwechslungen nachweisen. Wiederum sind die Versuchspersonen der RM – Gruppe auch bei einer ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen sehr gut bzw. besser als die der ESM – Gruppe in der Lage, zwischen den Serien und Quellen zu differenzieren, was sich in einem signifikanten Ergebnis zeigt. Insgesamt und innerhalb der Serien ergaben sich wieder keine statistisch bedeutsamen Unterschiede zwischen den Versuchsgruppen.
- c-c-a) Im statistischen Ergebnis gleichen die folgenden Befunde für die Serien 'PKW-Fahrrad' zu NT2 (nach acht Wochen) exakt den Befunden zu T1, also ebenfalls einer ersten Befragung – nach einer Woche. Es ist lediglich in der Richtungstendenz der schematypischen Intrusionen ein Unterschied zu verzeichnen.
Bezüglich korrekt berichteter Details sowie Quellenverwechslungen waren keinerlei signifikante Unterschiede zwischen den Gruppen zu verzeichnen.
- c-c-b) Die Jugendlichen, die sich die zweite Unfall-Serie 'PKW-Fahrrad' selbst ausdenken sollten begingen hierbei im Interview immens mehr Fehler insgesamt sowie im Bereich allgemeiner Erinnerungsfehler und konfabulierten auch deutlich mehr als diejenigen Jugendlichen, die eine Dia-Serie 'PKW-Fahrrad' erinnern sollten.

c-c-c) Auch bei einer ersten Befragung nach acht Wochen (NT2) zu den Serien 'PKW-Fahrrad' ergab sich hinsichtlich des Vorkommens schematypischer Intrusionen ein deutlicher Unterschied zwischen den Versuchsgruppen, allerdings mit dem Unterschied dass im Gegensatz zu T1 und T2 nun die RM – Gruppe mehr Intrusionen vorbringt als die ESM – Gruppe.

4.3.1.2 Ergebnisse der Veränderung der Gedächtnisleistung über die Zeit (T1-T2)

Die nachfolgenden Angaben [a-...] beziehen sich auf die externale 'source-monitoring' - Gruppe, die Angaben [b-...] auf die 'reality-monitoring' - Gruppe:

- a-a-a) Bezogen auf beide Dia-Serien ergaben sich hier keinerlei signifikante Veränderungen über die Zeit hinweg weder bezüglich der Gesamtzahl der korrekt berichteten Details oder im freien Bericht noch auf Nachfrage.
- a-a-b) Dagegen werden für beide Dia-Serien zusammen zu T2 hochsignifikant mehr Fehler insgesamt begangen als zu T1.
- a-a-c) Hinsichtlich der Arten von Quellenverwechslungen (insgesamt, innerhalb und zwischen den Dia-Serien) ergeben sich über beide Dia-Serien hinweg keine wesentlichen Veränderungen zwischen den beiden Befragungszeitpunkten.
- a-a-d) Gleiches gilt für schematypische Intrusionen sowie auch Konfabulationen. Zum zweiten Befragungszeitpunkt T2 unterlaufen den Versuchspersonen tendenziell weniger Intrusionen, dafür allerdings etwas mehr Konfabulationen als zu T1.
- a-a-e) Analog zur Gesamtfehlerzahl nimmt auch die Rate der allgemeinen Erinnerungsfehler von T1 nach T2 deutlich und statistisch hochsignifikant zu.

- a-b) Für die Dia-Serie 'PKW-PKW' ließen sich weder hinsichtlich der korrekt berichteten Details noch der Fehler und Fehlerarten wesentlich Veränderungen zwischen den Befragungszeitpunkten beobachten. Für keine der Variablen fand sich ein signifikantes Ergebnis.
- a-b-a) Tendenziell war für folgende Variablen ein leichter Anstieg in der Ausprägung des jeweiligen Mittelwerts vom ersten zum zweiten Befragungszeitpunkt zu verzeichnen: korrekte Details 'gesamt' und im 'free recall', Fehler 'gesamt', Quellenverwechslungen 'gesamt' und innerhalb der Dia-Serie, Konfabulationen und allgemeine Erinnerungsfehler

- a-b-b) Korrekte Details auf Nachfrage, Quellenverwechslungen zwischen den Dia-Serien sowie schematypische Intrusionen verzeichneten dagegen tendenziell zum zweiten Befragungszeitpunkt eher geringere Ausprägungen als zum ersten.
- a-c-a) Bezüglich der Dia-Serie 'PKW-Fahrrad' werden in dieser Versuchsgruppe zum zweiten Befragungszeitpunkt T2 sowohl deutlich mehr Fehler insgesamt als auch allgemeine Erinnerungsfehler begangen als zu T1.
- a-c-b) Hinsichtlich der korrekt berichteten Details sowie der übrigen Fehlerarten finden sich in dieser zweiten Dia-Serie keine bedeutsamen Unterschiede zwischen erstem und zweitem Befragungszeitpunkt.
- a-c-c) Leichte tendenzielle Veränderungen in Form eines Anstiegs bzw. Abfalls von T1 nach T2 konnten für korrekte Details gesamt und im freien Bericht, schematypische Intrusionen und Konfabulationen bzw. korrekte Details 'auf Nachfrage' beobachtet werden. Sämtliche Formen von Quellenverwechslungen zeigten auch keinerlei tendenzielle Veränderungen zwischen den Befragungszeitpunkten.
- b-a-a) Im Gegensatz zur ESM – Gruppe berichten die Versuchspersonen der RM – Gruppe zum zweiten Befragungszeitpunkt über beide Unfall-Serien hinweg signifikant weniger korrekte Details insgesamt und auch auf Nachfrage als zum ersten. Zu T2 werden im freien Bericht zwar etwas mehr korrekte Details berichtet als zu T1, jedoch ist dies statistisch unbedeutend.
- b-a-b) Wie auch die Jugendlichen der Versuchsgruppe 'external-external' machen die Jugendlichen der Gruppe 'external-internal' zum zweiten Befragungszeitpunkt T2 nicht nur deutlich mehr Fehler insgesamt sondern auch deutlich mehr allgemeine Erinnerungsfehler als zum ersten Befragungszeitpunkt T1 – bezogen auf beide Unfall-Serien.
- b-a-c) Die Veränderungen zwischen den Zeitpunkten bezüglich der Quellenverwechslungen erweisen sich auch für diese Versuchsbedingung erneut als dermaßen gering, dass kein statistisch bedeutsames Ergebnis festgehalten werden kann.
- b-a-d) Eben solches zeigt sich für das Auftreten schematypischer Intrusionen zu den beiden Zeitpunkten. Es ist lediglich ein leichter Abfall von T1 nach T2 erkennbar.
- b-a-e) Konfabulationen treten jedoch zum zweiten Befragungszeitpunkt signifikant häufiger auf als zum ersten, betrachtet man beide Unfall-Serien zusammen.
- b-b-a) Bei separater Betrachtung der Dia-Serie 'PKW-PKW' ergaben sich für die korrekt berichteten Details keine wesentlichen Veränderungen über die Zeit hinweg. Tenden-

- ziell zeigten sich hier insgesamt und im freien Bericht zu T2 jedoch höhere Werte als zu T1.
- b-b-b) Im Gegensatz zur externalen 'source-monitoring'- Gruppe fanden sich hier für die Dia-Serie 'PKW-PKW' zum zweiten Befragungszeitpunkt T2 sowohl wesentlich höhere Gesamtfehlerzahlen als auch wesentlich mehr allgemeine Erinnerungsfehler im Vergleich zu T1.
 - b-b-c) Was die Arten von Quellenverwechslungen aber auch vorgebrachte konfabulierte Inhalte betrifft, konnten wiederum kaum bzw. keine signifikanten Veränderungen zwischen den Befragungszeitpunkten festgestellt werden (Dia-Serie 'PKW-PKW').
 - b-b-d) Zwar zeigt sich hinsichtlich schematypischer Intrusionen ein signifikantes Ergebnis, jedoch kommen zu T2 weniger Intrusionen vor als zu T1.
-
- b-c-a) Im Gegensatz zur Dia-Serie 'PKW-Fahrrad' der ESM – Gruppe berichten die Jugendlichen der RM – Gruppe zum zweiten Befragungszeitpunkt deutlich weniger korrekte Details (gesamt und auf Nachfrage) ihrer selbst erfundenen Unfall-Serie 'PKW-Fahrrad' als zum ersten Befragungszeitpunkt. Im freien Bericht sind dies tendenziell auch weniger, jedoch statistisch nicht bedeutsam.
 - b-c-b) Allerdings zur Dia-Serie 'PKW-Fahrrad' der ESM – Gruppe analoge Ergebnisse ließen sich hinsichtlich der Gesamtfehlerzahl sowie der Anzahl der allgemeinen Erinnerungsfehler in der selbst erfundenen Unfall-Serie 'PKW-Fahrrad' feststellen. In beiden Fällen stieg die Fehlerzahl von T1 zu T2 immens und höchstsignifikant an.
 - b-c-c) Erneut veränderten sich Quellenverwechslungen und auch schematypische Intrusionen bei Betrachtung der selbst erfundenen Unfall-Serie 'PKW-Fahrrad' über die Zeit fast nicht und blieben im Ergebnis nicht-signifikant.
 - b-c-d) Allerdings konfabulierten die Jugendlichen zu T2 signifikant mehr als zu T1, sollten sie sich eine der Dia-Serie 'PKW-Fahrrad' analoge Unfall-Serie selbst ausdenken.

4.3.1.3 Ergebnisse der Gedächtnisleistungen im Vergleich der einzelnen Befragungszeitpunkte zueinander

a) Befragungszeitpunkt T1 vs. NT2

Die nachfolgenden Befunde [a-a...]) beziehen sich auf die externe 'source-monitoring' - Gruppe:

- a-a-a-a) Beim Vergleich der korrekten Details berichten die Jugendlichen bezogen auf beide Dia-Serien insgesamt wie auch im 'free recall' nach einer Woche zu T1 deutlich mehr Details als nach acht Wochen zu NT2. Auf Nachfrage sind es tendenziell zwar auch mehr, jedoch ist dieser Unterschied statistisch nicht-signifikant.
- a-a-a-b) Wie zu vermuten war, ist die Gesamtfehlerzahl sowie die Rate der allgemeinen Erinnerungsfehler bei Befragung nach acht Wochen (NT2) zwar höher als nach einer Woche (T1), allerdings erweisen sich beide Unterschiede statistisch gesehen als unbedeutend – bezogen auf beide Dia-Serien.
- a-a-a-c) Ebenso nicht-signifikant zeigt sich ein weiteres Mal der Vergleich der begangenen Quellenverwechslungen. Auch hier lassen sich über beide Dia-Serien hinweg keine wesentlichen Veränderungen zwischen T1 und NT2 nachweisen.
- a-a-a-d) Hochsignifikante Veränderungen ergaben sich jedoch sowohl hinsichtlich schematypischer Intrusionen als auch bezüglich Konfabulationen - allerdings in verschiedene Richtungen:
Während die Versuchspersonen bei Abfrage zu T1 in beiden präsentierten Dia-Serien mehr schematypische Intrusionen vorbrachten als zu NT2, wurde zu NT2 deutlich mehr konfabuliert als zu T1.
- a-a-b) Wie schon beim Vergleich der Gedächtnisleistungen über die Zeit (T1-T2) ergibt sich bei isolierter Betrachtung der Dia-Serie 'PKW-PKW' für die ESM – Gruppe hier ebenfalls für keine der Variablen eine signifikante Veränderung zwischen den Befragungszeitpunkten T1 und NT2.
Es lassen sich hier auch nur eventuelle Richtungstendenzen beobachten. Zu NT2 ergaben sich für die Gesamtfehlerzahl, Quellenverwechslungen innerhalb der Dia-Serie, Konfabulationen sowie allgemeine Erinnerungsfehler höhere Werte als zu T1, umgekehrt gilt dies für korrekte Details 'gesamt', im 'free recall' und 'auf Nachfrage', Quellenverwechslungen insgesamt und zwischen den Serien sowie schematypische Intrusionen.
- a-a-c-a) Völlig anders als beim Vergleich T1-T2 über die Zeit lassen sich bei Betrachtung der Dia-Serie 'PKW-Fahrrad' für sich allein für die Versuchspersonen der Bedingung 'external-external' sehr deutliche Unterschiede zwischen den Befragungszeitpunkten T1 und NT2 aufzeigen. Hier berichten die Jugendlichen insgesamt, im 'free recall' und auch auf Nachfrage bei einer ersten Befragung nach acht Wochen jeweils deutlich weniger korrekte Details als bei einer ersten Befragung nach einer Woche.

- a-a-c-b) Im Gegensatz zum Vergleich T1-T2 der ESM – Gruppe diesbezüglich gestalten sich die Befunde hinsichtlich der Gesamtfehlerzahl sowie der Zahl der allgemeinen Erinnerungsfehler. Beide Fehlerraten unterscheiden sich zu den Befragungszeitpunkten T1 und NT2 kaum und nicht wesentlich voneinander. Gleiches gilt innerhalb der Dia-Serie 'PKW-Fahrrad' für das Vorkommen von Quellenverwechslungen und Konfabulationen.
- a-a-c-c) Ein signifikanter Unterschied zwischen T1 und NT2 findet sich lediglich innerhalb der schematypischen Intrusionen; jedoch ist hier zu bemerken, dass diese für die Dia-Serie 'PKW-Fahrrad' nach acht Wochen weniger häufig vorkommen als nach einer Woche.

Die nachfolgenden Befunde [a-b-...] beziehen sich auf die 'reality-monitoring' - Gruppe

- a-b-a-a) Anders als beim Vergleich T1-NT2 für die ESM – Gruppe berichten die Versuchspersonen der RM - Gruppe über beide Unfall-Serien hinweg – zwar tendenziell zu NT2 ebenfalls weniger als zu T1 - signifikant allerdings nicht weniger korrekte Details insgesamt, im freien Bericht und auch auf Nachfrage hin.
- a-b-a-b) Wiederum erwartungsgemäß ergibt sich das Bild für die insgesamt begangenen sowie die allgemeinen Erinnerungs-Fehler. Auch die RM – Gruppe macht bei einer ersten Befragung nach acht Wochen wesentlich mehr dieser Fehler als bei einer ersten Befragung, die schon nach einer Woche stattfindet – bezogen auf beide Unfall-Serien. Gleiches gilt auch erneut für die Fehlerart 'schematypische Intrusionen'.
- a-b-a-c) Kaum feststellbare, nicht-signifikante Unterschiede zwischen T1 und NT2 sind bezüglich der Quellenverwechslung insgesamt, innerhalb und zwischen den Serien sowie konfabulierten Inhalten beider Unfall-Serien zu bemerken.
- a-b-b-a) Bei wiederum isolierter Betrachtung der Dia-Serie 'PKW-PKW' macht es für die Jugendlichen der Versuchsbedingung 'external-internal' offensichtlich keinen Unterschied hinsichtlich des Erinnerns korrekter Details, ob eine erste Befragung bereits nach einer Woche (T1) oder erst nach ca. acht Wochen (NT2) stattfindet. Hier finden sich keinerlei signifikante Unterschiede, sei es hinsichtlich der Gesamtzahl der erinnerten Details oder im freien Bericht oder auf Nachfrage.
- a-b-b-b) Dem gegenüber scheinen auch hier wieder die Fehlerraten entscheidend zu sein. Sowohl in der Gesamtfehlerzahl als auch was allgemeine Erinnerungsfehler angeht werden in der Dia-Serie 'PKW-PKW' zu NT2 deutlich und höchstsignifikant mehr Fehler begangen als zu T1.

- a-b-b-c) Keinerlei signifikante Unterschiede sind für die Dia-Serie 'PKW-PKW' hinsichtlich aller Formen von Quellenverwechslungen sowie Konfabulationen zu konstatieren. Tendenziell sind die diesbezüglichen Fehlerwerte zu NT2 jedoch höher als zu T1.
- a-b-b-d) Schematypische Intrusionen kommen für diese Dia-Serie allerdings nach acht Wochen signifikant häufiger vor als nach einer Woche.

- a-b-c-a) Betrachtet man dagegen die selbst erfundene Unfall-Serie per-se, kann hinsichtlich der korrekt berichteten Details in allen Abstufungen (gesamt, im freien Bericht und auf Nachfrage) festgestellt werden, dass diejenigen Jugendlichen, die nach einer Woche erstmals befragt werden deutlich mehr Details korrekt erinnern als diejenigen, die nach ca. acht Wochen erstmals befragt wurden.
- a-b-c-b) Analog dazu unterlaufen den zuletzt genannten auch wesentlich (höchstsignifikant) mehr Fehler insgesamt wie auch im Bereich der allgemeinen Erinnerungsfehler.
- a-b-c-c) Bezüglich der restlichen Fehlerarten (Quellenverwechslungen, Intrusionen und Konfabulationen) lassen sich nur tendenzielle Veränderungen in Form von zu NT2 höheren Werten als zu T1 beobachten. Für ein statistisch relevantes Ergebnis sind diese Unterschiede zu gering oder beruhen lediglich auf Zufall.

b) Befragungszeitpunkt T2 vs. NT2

Die nachfolgenden Befunde [b-a...]) beziehen sich auf die externale 'source-monitoring' - Gruppe:

- b-a-a-a) Versuchspersonen der ESM – Gruppe scheinen hinsichtlich beider Dia-Serien, wenn sie nach acht Wochen bereits ein zweites Mal befragt werden (T2), wesentlich besser und deutlich mehr Details korrekt erinnern zu können als wenn sie nach dem selben Zeitraum ein erstes Mal interviewt wurden (NT2). Dies gilt auch statistisch für die Gesamtzahl als auch für die im freien Bericht erinnerten korrekten Details. Tendenziell gilt dies nur für auf Nachfrage erinnerte Details.
- b-a-a-b) Dieses klare Bild kann in diesem Zusammenhang offenbar nicht auf die Gesamtfehlerrate und die Zahl der allgemeinen Erinnerungsfehler übertragen werden. Hier macht es demzufolge keinen Unterschied, ob nach acht Wochen eine erste oder schon eine zweite Befragung stattfindet. Zwischen diesen Befragungszeitpunkten lassen sich keine signifikanten Unterschiede feststellen – zumindest nicht bei gleichzeitiger Betrachtung beider Dia-Serien. Ebenso wenig finden sich bedeutsa-

- me Unterschiede zwischen T2 und NT2 hinsichtlich von Quellenverwechslungen und auch Konfabulationen.
- b-a-a-c) Bemerkenswerterweise kommen schematypische Intrusionen bei wiederholter Befragung nach acht Wochen deutlich häufiger vor als bei einmaliger Befragung ebenfalls nach acht Wochen.
- b-a-b) Hinsichtlich getrennter Analyse der Dia-Serien ergab sich in der externalen 'source-monitoring' - Gruppe für die Dia-Serie 'PKW-PKW' außer bei den korrekt berichteten Details im freien Bericht für keine der anderen Variablen ein statistisch bedeutsamer Unterschied zwischen T2 und NT2. Innerhalb dieser Dia-Serie scheint es offensichtlich egal zu sein, ob die Versuchspersonen nach acht Wochen schon einmal zuvor befragt wurden oder nicht. Bezüglich der korrekt erinnerten Details lässt sich zu T2 tendenziell eine höhere Erinnerungsrate konstatieren als zu NT2. Über alle Arten von Fehlern lassen sich kaum fundierte Aussagen treffen.
- b-a-c-a) Für die Dia-Serie 'PKW-Fahrrad' zeigt sich in dieser Versuchsgruppe hinsichtlich der korrekt berichteten Details insgesamt und im freien Bericht ein anderes Bild. Hypothesenkonform erinnern die Jugendlichen bei einer zweiten Befragung (T2) nach acht Wochen wesentlich mehr Details als bei einer ersten Befragung nach acht Wochen (NT2). Tendenziell, jedoch statistisch nicht belegt, scheint dies auch für die auf Nachfrage korrekt erinnerten Details zu gelten.
- b-a-c-b) Auch als nicht statistisch bedeutsam erweisen sich die Unterschiede zwischen diesen beiden Befragungszeitpunkten bezogen auf die zweite präsentierte Dia-Serie 'PKW-Fahrrad' für die Gesamtfehlerzahl, die Zahl der allgemeinen Erinnerungsfehler, der Arten von Quellenverwechslungen und Konfabulationen.
- b-a-c-c) Lediglich hinsichtlich schematypischer Intrusionen scheint hier von Bedeutung zu sein, ob die Versuchspersonen nach acht Wochen zum ersten Mal oder zum zweiten Mal befragt wurden. Bei einer zweiten Befragung finden sich hochsignifikant mehr Intrusionen als bei einer ersten.

Die nachfolgenden Befunde [b-b-...] beziehen sich auf die 'reality-monitoring' - Gruppe:

- b-b-a-a) Im Gegensatz den Versuchspersonen der Bedingung 'external-external' erinnern die Jugendlichen dieser Versuchsbedingung bezüglich beider Unfall-Serien offenbar genau so gut oder schlecht korrekte Details zu T2 wie zu NT2, sei es insgesamt, im 'free recall' oder auf Nachfrage. Statistisch gesehen ergaben sich diesbezüglich zwischen den Befragungszeitpunkten keine Unterschiede.

- b-b-a-b) Ebenso umgekehrt zur ESM – Gruppe verhält es sich bezüglich der Unterschiede zwischen T2 und NT2 für die insgesamt gemachten Fehler sowie die allgemeinen Erinnerungsfehler. Bei erstmaliger Befragung nach acht Wochen treten hier für beide Fehlervariablen – betrachtet man beide Unfall-Serien simultan – sehr viel mehr Fehler auf als zu T2 bei bereits vorangegangener Befragung.
- b-b-a-c) Wie schon in den voranstehenden Darstellungen auch ergaben sich auch hier beim Vergleich der begangenen Quellenverwechslungen und dem Auftreten von Konfabulationen keine relevanten Unterschiede zwischen den Befragungszeitpunkten.
- b-b-a-d) Von Bedeutung sind jedoch auch hier wieder einmal schematypische Intrusionen, welche in dieser Versuchsbedingung über beide Unfall-Serien hinweg bei einer ersten Befragung nach acht Wochen (NT2) viermal häufiger genannt werden als bei einer wiederholten Befragung nach acht Wochen (T2).

- b-b-b-a) Bei wiederum getrennter Betrachtung der Unfall-Serien ergab sich zunächst für die Dia-Serie 'PKW-PKW' lediglich für die Gesamtfehlerzahl sowie die schematypischen Intrusionen ein signifikanter Unterschied zwischen T2 und NT2. In beiden Fällen traten hier nach acht Wochen zu NT2 deutlich mehr Fehler auf als zu T2 mit einer vorausgegangenen Befragung.
- b-b-b-b) Hinsichtlich aller übrigen Fehlerarten sowie der korrekt berichteten Details scheint es in der Bedingung 'external-internal' im statistischen Ergebnis nicht so sehr von Bedeutung zu sein, ob nach acht Wochen eine erste oder eine zweite Befragung stattfindet. Tendenziell ergaben sich jedoch hypotesenkonforme Ergebnisse in Form von zu NT2 höheren Fehlerwerten und zu T2 höheren Werten bezüglich der korrekt berichteten Details.

- b-b-c-a) Innerhalb der selbst erfundenen Unfall-Serie fanden sich nur bei drei der zehn abhängigen Variablen (hoch)signifikante Unterschiede zwischen T2 und NT2, nämlich hinsichtlich der Gesamtfehlerzahl sowie der Rate der allgemeinen Erinnerungsfehler (wie auch in der ESM – Gruppe) als auch hinsichtlich der schematypischen Intrusionen. Diejenigen Versuchspersonen, welche nach acht Wochen zum ersten Mal interviewt wurden, machten dann jeweils deutlich mehr Fehler in den genannten Formen als diejenigen, die sich zu diesem Zeitpunkt schon einer Befragung unterzogen hatten.
- b-b-c-b) Für die übrigen Fehlerarten sowie die korrekt berichteten Details ließen sich jeweils nur unwesentliche, statistisch irrelevante Unterschiede zwischen T2 und NT2 nachweisen. Diesbezüglich scheint für die Gruppe 'external-internal' die Anzahl der

Befragungen nach acht Wochen in diesem Rahmen von geringerer Bedeutung zu sein

4.3.1.4 Ergebnisse zu den Haupt- und Interaktionseffekten

- a) Weder für die Gesamtzahl der korrekt berichteten Details noch für die korrekt berichteten Details im freien Bericht und auch nicht auf Nachfrage ließen sich signifikante Haupteffekte der Faktoren 'Versuchsbedingung' bzw. 'Befragungszeitpunkt' nachweisen. Ebenso fand sich diesbezüglich keine Wechselwirkung zwischen den beiden Faktoren.
- b) Dagegen konnten hinsichtlich der Gesamtfehlerzahl sowie auch der Fehlerzahl der allgemeinen Erinnerungsfehler jeweils sowohl höchstsignifikante Haupteffekte 'Gruppe' wie auch 'Befragungszeitpunkt' als auch eine höchstsignifikante Interaktion zwischen diesen festgestellt werden.
- c) Mit Ausnahme eines signifikanten Haupteffekts der Gruppenzugehörigkeit bezüglich der Quellenverwechslungen zwischen den Serien konnten für die Arten von Quellenverwechslungen (insgesamt, innerhalb und zwischen den Serien) weder signifikante Haupteffekte noch statistisch bedeutsame Zusammenhänge zwischen der Versuchsbedingung und dem Befragungszeitpunkt gefunden werden.
- d) Schematypische Intrusionen trennen in Form eines hochsignifikanten Haupteffekts zwar zwischen den Versuchsgruppen an sich, bezüglich des Befragungszeitpunktes ergab sich jedoch kein signifikantes Ergebnis.
In Abhängigkeit vom Befragungszeitpunkt ließ sich diesbezüglich jedoch kein Unterschied zwischen den Versuchsgruppen nachweisen.
- e) Ein wiederum statistisch höchstbedeutsamer Unterschied zwischen den Versuchsbedingungen ergab sich als Haupteffekt hinsichtlich des Auftretens von konfabulierten Inhalten.
Ebenso statistisch relevant - weil signifikant - zeigt sich das Ergebnis beim Vergleich der Befragungszeitpunkte. Die Befragungszeitpunkte unterscheiden sich im Hinblick auf Konfabulationen in Form eines Haupteffekts deutlich voneinander.
Allerdings konnte zwischen den beiden Faktoren keine gegenseitige Abhängigkeit oder Wechselwirkung beobachtet werden.

4.3.2 Glaubhaftigkeit von Zeugenaussagen

Im Folgenden werden zunächst die Ergebnisse der Validierungsuntersuchung der Aussagemerkmale (Realkennzeichen) im Einzelnen sowie der fünf Merkmalskategorien und aller Realkennzeichen insgesamt dargestellt. Anschließend folgt eine kurze Zusammenfassung der Ergebnisse zur Glaubhaftigkeitseinschätzung der Aussagen im Rahmen des Ratings bzw. der Auswertung.

4.3.2.1 Ergebnisse zu den einzelnen Aussagemerkmalen

a) Folgende Aussagemerkmale wiesen sowohl zu T1 als auch zu T2 **in erlebnis-basierten Aussagen deutlich höhere Merkmalsausprägungen auf als in nicht erlebnis-basierten** und stellen somit sehr gute, vom Zeitpunkt unabhängige Diskriminatoren dar:

- Logische Konsistenz
- Widerspruchsfreie, ungeordnete Darstellung
- Kontextuelle Einbettung
- Schilderung nonverbaler Aktionen
- Nebensächliche Details
- Schilderung eigener psychischer oder physischer Vorgänge
- Einwände gegen die Richtigkeit der eigenen Aussage
- Deliktspezifische Aussageelemente

Nur zum ersten Befragungszeitpunkt T1 waren lediglich bei der 'Schilderung von Komplikationen im Handlungsverlauf' die Mittelwerte in den wahren Geschichten wesentlich höher als in den erfundenen.

Nur zum zweiten Befragungszeitpunkt T2 konnten bei der 'Wiedergabe von Gesprächen' sowie der 'Schilderung ausgefallener Details' in den erlebnis-fundierte Aussagen bedeutsam und relevant höhere Merkmalsausprägungen festgestellt werden als in den nicht erlebnis-fundierte.

Für keines der Glaubhaftigkeitskriterien war zu keinem der beiden Befragungszeitpunkte in den erfundenen Geschichten eine wesentlich oder statistisch bedeutsame höhere Merkmalsausprägung zu beobachten als in den wahren.

b) Ein **deutlicher Abfall in der Merkmalsausprägung von T1 nach T2** konnte für beide Geschichten lediglich hinsichtlich der 'Schilderung nonverbaler Aktionen' sowie der 'Schilderung eigener psychischer oder physischer Vorgänge' nachgewiesen werden.

Für erlebnis-basierte Aussagen fand sich dieser Rückgang in der Merkmalsausprägung hinsichtlich 'nebensächlicher Details', 'spontaner Verbesserungen der eigenen Aussage' sowie 'Selbstbelastungen'.

Für nicht erlebnis-basierte Aussagen waren bezüglich der 'Wiedergabe von Gesprächen', der 'Schilderung von Komplikationen im Handlungsverlauf' sowie der Schilderung 'ausgefallener Details' die Mittelwerte zum zweiten Befragungszeitpunkt T2 geringer als zum ersten Befragungszeitpunkt T1.

Keines der Realkennzeichen wies weder für wahre noch für erfundene Geschichten - wenn überhaupt - einen wesentlichen Anstieg in der Merkmalsausprägung von T1 nach T2 auf.

c) Folgende Aussagemerkmale **trennen** zu beiden Befragungszeitpunkten T1 und T2 **zwischen Aussagen mit Erlebnishintergrund und Aussagen ohne Erlebnishintergrund**:

- Logische Konsistenz
- Ungeordnete, widerspruchsfreie Darstellung
- Quantitativer Detailreichtum
- Kontextuelle Einbettung
- Schilderung nonverbaler Aktionen
- Nebensächliche Details
- Schilderung eigener psychischer oder physischer Vorgänge
- Einwände gegen die Richtigkeit der eigenen Aussage
- Deliktspezifische Aussageelemente

Bei Betrachtung des ersten Befragungszeitpunktes T1 für sich ergibt sich ein deutlicher Unterschied zwischen wahren und erfundenen Geschichten nur hinsichtlich 'spontaner Verbesserungen der eigenen Aussage'.

Allein zum zweiten Befragungszeitpunkt T2 trennen die Realkennzeichen 'Wiedergabe von Gesprächen', 'Schilderung von Komplikationen im Handlungsverlauf' sowie die 'Schilderung ausgefallener Details'.

4.3.2.2 Ergebnisse zu den Merkmalskategorien und zu allen Aussagemerkmalen als Gesamtinstrument

- a) Bezüglich der ersten und diagnostisch sehr bedeutsamen Merkmalskategorie 'Allgemeine Merkmale' konnten zu beiden Befragungszeitpunkten T1 und T2 in den er-

lebnisfundierten Aussagen deutlich höhere Ausprägungen festgestellt werden als in den nicht erlebnis-fundierten.

Allerdings finden hier weder für wahre noch erfundene Geschichten bedeutsame Veränderungen über die Zeit hinweg von T1 nach T2 statt.

- b) Für den Bereich der 'Speziellen Inhalte' fanden sich zum einen deutlich höhere Merkmalsausprägungen in den wahren Geschichten gegenüber den erfundenen (sowohl zu T1 als auch zu T2), zum anderen zum zweiten Befragungszeitpunkt deutlich geringere als zum ersten Befragungszeitpunkt (sowohl für wahre als auch für erfundene Geschichten).
- c) Exakt analog zu den 'Speziellen Inhalten' gestalten sich die Befunde hinsichtlich der 'Inhaltlichen Besonderheiten'.
- d) Innerhalb des Merkmalskomplexes 'Motivationsbezogene Inhalte' zeigte sich lediglich zum ersten Befragungszeitpunkt T1 ein deutlicher höherer Wert in der Merkmalsausprägung in den erlebnis-basierten Aussagen gegenüber denjenigen ohne Erlebnishintergrund.
Wie auch hinsichtlich der 'Allgemeinen Merkmale' verändern sich die Merkmalsausprägungen weder für wahre noch für erfundene Geschichten nicht wesentlich vom ersten zum zweiten Befragungszeitpunkt.
- e) Wiederum analog - zu den Ergebnissen der 'Allgemeinen Merkmale' - können die Befunde bezüglich des Vorkommens 'Deliktspezifischer Inhalte' zusammengefasst werden. Zwar sind die jeweiligen Merkmalsausprägungen aufgrund der Tatsache einer Simulationsstudie und der Wahl der Themen recht gering, jedoch sind zu beiden Befragungszeitpunkten im Mittel die Werte der erlebnis-fundierten Aussagen höher als die der nicht erlebnis-fundierten.
Eine bedeutsame Veränderungen über die Befragungszeitpunkte hinweg findet sich allerdings auch hier nicht – weder für wahre noch für erfundene Geschichten.
- f) Für keine der Merkmalskategorien konnte eine wechselseitige Abhängigkeit zwischen Erlebnissbasis und Befragungszeitpunkt (Interaktion) beobachtet werden.
- g) Betrachtet man alle in die Untersuchung einbezogenen Aussagemerkmale quasi als ein Gesamtinstrument zur Differenzierung der Glaubhaftigkeit bzw. zur Überprüfung eines gegebenenfalls vorhandenen Erlebnishintergrundes, so wird deutlich, dass in erlebnis-

fundierten Aussagen die Realkennzeichen deutlich stärker ausgeprägt sind als in den Aussagen ohne Erlebnishintergrund. Dies gilt sowohl über beide Befragungszeitpunkte hinweg betrachtet als auch für beide Zeitpunkte einzeln.

Insgesamt findet sich sowohl in wahren als auch in erfundenen Geschichten zum zweiten Befragungszeitpunkt eine deutlich geringere Ausprägung der Summe aller Merkmale gegenüber dem ersten.

Eine Wechselwirkung zwischen dem 'Wahrheitsstatus der Geschichten' und dem 'Befragungszeitpunkt' zeigt sich nicht.

4.3.2.3 Ergebnisse zur Glaubhaftigkeitseinschätzung

- a) Über beide Befragungszeitpunkte hinweg konnten 89 % der Aussagen richtig identifiziert bzw. klassifiziert werden, 10 % wurden falsch eingestuft.

Getrennt nach Befragungszeitpunkten ergeben sich sehr ähnliche Ergebnisse:

- b) Zum ersten Befragungszeitpunkt T1 ergab sich eine positive Trefferquote von 88 %, d.h. es wurden 44 von 50 Aussagen richtig gemäß ihres tatsächlichen Wahrheitsstatus identifiziert. 12 % der Aussagen wurden falsch eingestuft.
- c) Zum zweiten Befragungszeitpunkt T2 steigt die Rate der korrekt erkannten Aussagen im Vergleich zu T1 um 2 % auf 90 % leicht an, während die negative Trefferquote der falsch klassifizierten Aussagen auf 8 % abfällt gegenüber 12 % zum ersten Befragungszeitpunkt.

Insgesamt kann eine recht gute Glaubhaftigkeitseinschätzung festgestellt werden, wobei dies tendenziell zu einem weiteren, später stattfindenden Befragungszeitpunkt besser bzw. leichter zu sein oder zu werden scheint. Die Rolle eines zweiten Befragungszeitpunktes scheint sich hier wiederum als wichtig zu erweisen und bleibt im Rahmen der Bedeutung und des Vorgehens in der forensischen Praxis und Begutachtung und auch für die vorliegende Arbeit zu diskutieren – insbesondere im Bereich und bezogen auf Personen mit Intelligenzminderung.

4.3.2.4 Ergebnisse zur Klassifikation

Die durchgeführte Diskriminanzanalyse beantwortet zufriedenstellend letztendlich die anfangs gestellten Grundfragen:

- (1) Unterscheiden sich die Gruppen signifikant voneinander hinsichtlich der Variablen?
- (2) Welche Variablen sind zur Unterscheidung zwischen den Gruppen geeignet bzw. ungeeignet?

Es ergibt sich zunächst, dass die jeweils ermittelten Diskriminanzfunktionen zu beiden Zeitpunkten sehr deutlich zwischen den Versuchsgruppen trennen können.

Zum ersten Befragungszeitpunkt T1 bleiben dann von den ursprünglich 19 erhobenen (bzw. 17 für die varianzanalytischen Verfahren) und tatsächlich in die Analyse einbezogenen Merkmalsvariablen acht diskriminatorisch bedeutsame Kriterien übrig, die für sich alleine schon völlig zur Varianzaufklärung beitragen. Durch diese acht verbleibenden Aussagemerkmale konnten letztendlich 96 % der ursprünglich gruppierten Fälle korrekt klassifiziert werden. Die Diskriminanzfunktion liegt also mit ihrer „Treffer-“ bzw. Klassifizierungsquote um 8 % höher als die Quote des Ratings (88 %).

Zum zweiten Befragungszeitpunkt T2 kristallisieren sich neun diskriminatorisch bedeutsame Variablen heraus, die für sich genommen in der Lage wären als Analyseinstrument zwischen erlebnisfundierten und nicht-erlebnisfundierten Aussagen zu unterscheiden. Diese neun Aussagemerkmale klassifizieren ebenfalls 96 % der ursprünglich durch die Rater gruppierten Fälle korrekt. Damit liegt die korrekte Klassifizierungsrate mit einer „Differenz“ von 6 % auch zum zweiten Befragungszeitpunkt T2 etwas höher gegenüber der Quote des Ratings (90 %).

Zu beiden Befragungszeitpunkten stellten sich hierbei vor allem die Aussagemerkmale aus dem Bereich der 'Allgemeinen Merkmale' als bedeutsamste Diskriminationskriterien heraus, insbesondere das Merkmal 'Logische Konsistenz', das zu beiden Befragungszeitpunkten an erster Stelle der Klassifikationsmatrix steht.

Als ebenso diskriminatorisch sehr bedeutsam erwiesen sich vor allem die 'Kontextuelle Einbettung' von Aussageelementen sowie die 'Schilderung nonverbaler Aktionen' (spezielle Inhalte) - sowohl zum ersten als auch zum zweiten Befragungszeitpunkt.

Aus dem Bereich der 'Inhaltlichen Besonderheiten' scheint der 'Schilderung eigener psychischer und physischer Vorgänge' - ebenfalls zu beiden Befragungszeitpunkten - ein hoher diskriminatorischer Wert zu zukommen.

5 Diskussion

Zunächst sollen die Befunde zur Erinnerungsleistung und Quellenidentifikation bzw. Quellenüberwachung diskutiert und näher beleuchtet werden (Kapitel 5.1).

Im Anschluss daran (Kapitel 5.2) folgt die Diskussion zu den Befunden der Validierungsuntersuchung der aussageimmanenten Qualitätsmerkmale (Glaubhaftigkeit von Zeugenaussagen).

Nach einer kritischen Betrachtung des wissenschaftlichen Erkenntniswerts, sonstigen Aspekten (Kapitel 5.3) der vorliegenden (Teil-) Untersuchung werden Ausblicke und Implikationen für die forensische Praxis und weitere Forschungsarbeiten aufgezeigt (Kapitel 5.4).

5.1 Diskussion zu den Befunden der Erinnerungsleistung und dem source- bzw. reality monitoring

In diesem Kapitel werden zunächst die Ergebnisse der Unterschiede, die sich in der Gedächtnisleistung der beiden untersuchten Gruppen zu den einzelnen Zeitpunkten (T1, T2 und NT2) ergaben, erläutert. Anschließend werden die Befunde der Veränderung der Gedächtnis- und Erinnerungsleistung über die Zeit (T1–T2), der beiden Befragungszeitpunkte (T1-NT2 sowie T2-NT2) im Vergleich und der untersuchten Interaktionseffekte begründet und diskutiert.

Grundsätzlich ist eine direkte Gegenüberstellung bzw. ein Vergleich der vorliegenden Befunde mit anderen Studien zu diesem Thema nicht oder nur begrenzt möglich, da in vielen anderen Untersuchungen mit Kontrollgruppen normalbegabter Kinder oder Erwachsener innerhalb der jeweiligen Studie gearbeitet wurde. Die Ergebnisse und Befunde der vorliegenden Untersuchung und ihrer Fragestellung stellen jedoch als Ganzes die Referenzuntersuchung zu drei parallel dazu am Lehrstuhl durchgeführten Diplomarbeiten dar bzw. umgekehrt. In zwei der drei Diplomarbeiten waren normalbegabte Schüler etwa desselben Alters die Versuchspersonen. Insofern können an dieser Stelle nur Aspekte innerhalb der Arbeit selbst diskutiert werden bzw. Vergleiche mit anderen Stichprobengruppen intelligenzgeminderter Personen aus anderen Untersuchungen. Ein Vergleich mit normal begabten Schülern und Jugendlichen ist daher nur indirekt mit den erwähnten Diplomarbeiten möglich. Eine Integration einer normal begabten Stichprobengruppe in diese Untersuchung hätte den Rahmen dieser noch mehr gesprengt.

5.1.1 Diskussion zu den Ergebnissen zwischen den Versuchsgruppen

Betrachtet man zunächst zum *ersten Befragungszeitpunkt T1* die Befunde über beide *Dia-* bzw. *Unfall-Serien* hinweg, so berichtet die 'reality-monitoring' - Gruppe zwar erkennbar mehr, jedoch nicht signifikant mehr korrekte Details insgesamt als die externe 'source-monitoring' - Gruppe, auf Nachfrage sind es allerdings signifikant mehr korrekte Details. Im freien Bericht sind die diesbezüglichen Werte zu T1 nahezu identisch.

Hinsichtlich der insgesamt begangenen Fehler konnte in der Gruppe 'external-internal' eine signifikant höhere Fehlerquote als in der Gruppe 'external-external' festgestellt werden, wohingegen sich bezüglich allgemeiner Erinnerungsfehler keine bedeutsamen Unterschiede feststellen ließen.

Hinsichtlich der Quellenverwechslungen insgesamt, innerhalb und zwischen den Serien sind die Unterschiede sowie die einzelnen Werte recht gering. Lediglich zwischen den Serien finden sich in der Versuchsbedingung 'external-external' signifikant mehr Quellenverwechslungen als in der Bedingung 'external-internal'.

Die Versuchspersonen der Bedingung, die zwischen zwei externalen Quellen unterscheiden müssen, begehen zu diesem ersten Befragungszeitpunkt signifikant mehr schematypische Intrusionen als die Jugendlichen der Bedingung 'external-internal'. Konfabulationen treten zu diesem ersten Befragungszeitpunkt jedoch sogar hochsignifikant mehr in der 'reality-monitoring' - Gruppe, also in der Versuchsgruppe, die zwischen einer externalen und einer internalen Quelle differenzieren sollten, auf.

Bei einzelner Betrachtung der *Dia-Serie 'PKW-PKW'* zum *ersten Befragungszeitpunkt T1* ist festzuhalten, dass die RM – Gruppe insgesamt sowie auf Nachfrage deutlich mehr korrekte Details erinnert und berichtet als die ESM – Gruppe. Im freien Bericht sind dies nur tendenziell mehr.

Hinsichtlich der insgesamt begangenen Fehler, der allgemeinen Erinnerungsfehler, der berichteten schematypischen Intrusionen sowie der Konfabulationen ließen sich für die *Dia-Serie 'PKW-PKW'* zwischen den Versuchsgruppen im Ergebnis keine signifikanten Unterschiede feststellen.

Wie auch über beide Serien hinweg betrachtet ergibt sich auch hier lediglich zwischen den Serien ein bedeutsamer Unterschied zwischen den Jugendlichen der Versuchsgruppen hinsichtlich aufgetretener Quellenverwechslungen, wiederum dergestalt, dass solche Fehler in der ESM – Gruppe deutlich häufiger zu finden sind als in der RM – Gruppe. Dort fanden sich überhaupt keine Quellenverwechslungen zwischen den Serien. Das Ausmaß der Quellenverwechslungen insgesamt sowie innerhalb der *Dia-Serie* war in beiden Versuchsgruppen erneut statistisch unbedeutend.

Völlig gegensätzlich verhalten sich die Beobachtungen und Ergebnisse zwischen den beiden Versuchsgruppen hinsichtlich *des ersten Befragungszeitpunktes für die Dia- bzw. Unfall-Serie 'PKW-Fahrrad'*.

Hier fanden sich weder hinsichtlich korrekt berichteter Details noch irgendeiner Art von Quellenverwechslungen nennenswerte Unterschiede oder signifikante Ergebnisse zwischen den Versuchsbedingungen 'external-external' und 'external-internal'.

Allerdings scheinen die Gedächtnisleistungen der Versuchspersonen der Gruppe RM hinsichtlich der übrigen Fehlerarten (allgemeine Erinnerungsfehler und Konfabulationen) sowie auch der Gesamtfehlerzahl weitaus störanfälliger zu sein als die der Gruppe ESM. Es zeigen sich jeweils sehr deutliche Unterschiede.

Im Ergebnis ebenfalls signifikant erweist sich der Befund der schematischen Intrusionen, jedoch kommen diese hier in der RM – Gruppe weniger oft vor als in der ESM – Gruppe.

Vorweg wurden zunächst auch zu einem *zweiten Befragungszeitpunkt T2* wiederum Vergleiche *über beide Dia- bzw. Unfall-Serien* angestellt. Wie zu T1 lassen sich auch hier hinsichtlich der insgesamt und auf Nachfrage korrekt berichteten Details keine signifikanten Unterschiede zwischen den beiden Versuchsgruppen feststellen. Dies gilt hier außerdem für die korrekten Details aus dem freien Bericht. Die Unterschiede zeigen sich hier geringer als zu T1.

Auch was die Gesamtfehlerzahl angeht, zeigt sich ein analoges Ergebnis zum ersten Befragungszeitpunkt. Den Jugendlichen der Versuchs-Gruppe 'external-internal' unterlaufen hier insgesamt sehr viel mehr Fehler als den Jugendlichen der Versuchs-Gruppe 'external-external'.

Wiederum finden sich keinerlei signifikante Veränderungen bezüglich der Quellenverwechslungen insgesamt, innerhalb oder zwischen den Serien.

Hinsichtlich des Vorkommens von schematischen Intrusionen sowie auch von Konfabulationen zeigen sich zu diesem zweiten Befragungszeitpunkt T2 jeweils analoge Befunde wie zum ersten Befragungszeitpunkt: Intrusionen kommen hier deutlich mehr in der externalen 'source-monitoring' - Gruppe vor als in der 'reality-monitoring' - Gruppe. Umgekehrt konfabulieren die Jugendlichen der Versuchsbedingung 'external-internal' deutlich mehr als die Jugendlichen der Vergleichsgruppe.

Allgemeine Erinnerungsfehler scheinen hier jedoch nach ca. sieben Wochen zu einem zweiten Befragungszeitpunkt im Gegensatz zum ersten an Bedeutung zu gewinnen. Die zu T1 zwar schon vorhandene Tendenz zeigt sich hier in einem auch statistisch höchst bedeutsamen höheren Fehlerwert der RM – Gruppe gegenüber der ESM - Gruppe.

Bei erneut getrennter Analyse der Dia-Serien scheint für die *Dia-Serie 'PKW-PKW'* hinsichtlich des Erinnerns korrekter Details auch nach acht Wochen (*zweiter Befragungszeitpunkt T2*) dasselbe zu gelten wie nach einer Woche zu T1. Insgesamt sowie auf Nachfrage berichten die Jugendlichen der RM – Gruppe wesentlich mehr Details als die der ESM – Gruppe. Im freien Bericht sind es wiederum auch nur tendenziell mehr.

Anders als zu T1 ist nun die Gesamtfehlerrate sowie die Anzahl im Bereich der allgemeinen Erinnerungsfehler in der Bedingung 'external-internal' deutlich höher als in der Bedingung 'external-external'.

Ebenso bestätigt bzw. wiederholt sich das Ergebnis (bei vergleichender Betrachtung von T1 und T2) bezüglich der Unterschiede der zwischen den Serien begangenen Quellenverwechslungen. Auch nach acht Wochen sind diese in der Gruppe RM nicht zu finden, jedoch wiederum schon in der Gruppe ESM, was sich in einem erneut signifikanten Ergebnis widerspiegelt.

Anhand sonstiger Quellenverwechslungen, Intrusionen und Konfabulationen können die beiden Versuchsgruppen wie auch zum ersten Befragungszeitpunkt nicht statistisch bedeutsam unterschieden werden.

Das zwar tendenziell schon von T1 bekannte Ergebnis zeigt sich nun *bei erneuter Befragung nach acht Wochen (T2)* auch statistisch zumindest was die insgesamt und im free recall berichteten Details angeht. Bezüglich der *Dia- bzw. Unfall-Serie 'PKW-Fahrrad'* erinnern und berichten die Versuchspersonen der RM – Gruppe signifikant weniger als diejenigen der ESM – Gruppe. Auf Nachfrage hin bleibt der Unterschied allerdings nur richtungsgemäß tendenziell erkennbar.

Zum zweiten Befragungszeitpunkt T2 können ebenfalls die statistischen Befunde hinsichtlich der Gesamtfehlerzahl, der Zahl der allgemeinen Erinnerungsfehler, der schematischen Intrusionen sowie der Konfabulationen auch exakt (d.h. auch bezüglich der Richtung der Mittelwertsunterschiede) nicht nur bestätigt und als wiederholt angesehen werden; die jeweiligen Unterschiede sind für die Serien 'PKW-Fahrrad' deutlich größer als zu T1 und zeigen sich in entsprechend hoch- und höchstsignifikanten Ergebnissen.

Hinsichtlich der Quellenverwechslungen ergibt sich ebenfalls wie zu T1 keinerlei signifikanter Unterschied zwischen den Gruppen.

In fast exakt analoger Weise zu einem Befragungszeitpunkt T2 gestalten sich die Ergebnisse *über beide Dia- bzw. Unfall-Serien* bezüglich der korrekt berichteten Details zu einem *Befragungszeitpunkt NT2*, ebenfalls nach ca. acht Wochen, der jedoch für eine Reihe der Versuchspersonen die erste und einzige Befragung darstellte. Auch zu diesem Befragungszeitpunkt scheinen korrekt berichtete Details hier weder insgesamt, im freien Bericht noch auf

Nachfrage eine geeignete Diskriminanzfunktion aufzuweisen. Es ergeben sich keine signifikanten Unterschiede zwischen den Versuchsbedingungen.

Dahingegen können die beiden Versuchsgruppen bezüglich der Gesamtfehlerzahl ein weiteres Mal sehr gut unterschieden werden. Wiederum finden sich hier in der 'reality monitoring' - Gruppe wesentlich mehr Fehler als in der externalen 'source-monitoring' - Gruppe. Der Unterschied ist zu diesem Befragungszeitpunkt im Vergleich zu den beiden anderen mit Abstand am größten. In der Gesamtbetrachtung lässt sich jeweils ein Anstieg der absoluten Gesamtfehlerzahl in der RM - Gruppe sowie auch des Unterschieds zwischen den Gruppen von T1 nach T2 nach NT2 konstatieren - vergleicht man die drei Befragungszeitpunkte diesbezüglich direkt miteinander.

In Analogie zu T2 begehen die Versuchspersonen der 'reality-monitoring' - Gruppe wesentlich mehr allgemeine Erinnerungsfehler als diejenigen der externalen 'source-monitoring' - Gruppe.

Ein weiteres Mal (wie auch zu T2) können die Versuchsgruppen anhand der Quellenverwechslungen (insgesamt, innerhalb und zwischen den Serien) nicht unterschieden werden, also unabhängig davon ob die Versuchspersonen nach acht Wochen ein erstes oder bereits ein zweites Mal befragt wurden.

Im Gegensatz zu den Befragungszeitpunkten T1 und T2 finden sich zu NT2 deutlich und signifikant mehr schematypische Intrusionen in der Versuchsbedingung 'external-internal' als in der Bedingung 'external-external'.

Wiederum und tendenziell mehr konfabuliert hingegen wird – wie zu T1 und T2 auch – in der Gruppe 'external-internal' gegenüber der Gruppe 'external-external'. Allerdings erweist sich der Unterschied hier statistisch gesehen als nicht-signifikant.

Bei einer *ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen (NT2)* ergaben sich nun sowohl für die Gesamtzahl als auch für die im freien Bericht sowie auf Nachfrage korrekt berichteten Details sehr deutliche Unterschiede zwischen ESM- und RM - Gruppe, wobei letztere sehr viel mehr Einzelheiten erinnert bzw. äußert.

Im Gegensatz zu T1 (nach einer Woche), jedoch identisch zu T2 (nach acht Wochen) begeht die Gruppe RM offensichtlich bei einer ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen wesentlich mehr Fehler insgesamt wie auch im Bereich der allgemeinen Erinnerungsfehler als die Gruppe ESM. Hier scheinen also offensichtlich sowohl die Häufigkeit der Befragungen wie auch der zeitliche Abstand zwischen diesen von Bedeutung für die Produktion von falsch erinnerten bzw. vergessenen Gedächtnisinhalten zu sein.

Wie auch schon zu T1 und auch zu T2 scheinen innerhalb der Dia-Serie 'PKW-PKW' schematypische Intrusionen sowie Konfabulationen auch zu diesem Befragungszeitpunkt keine entscheidende Rolle bei der Differenzierung zwischen den beiden Versuchsbedingungen zu

spielen. Die genannten Fehlerarten kommen in beiden Gruppen in etwa im selben Ausmaß vor, es konnten somit auch hier keine signifikanten Unterschiede festgestellt werden.

Ebenfalls ein statistisch identischer Befund ließ sich erneut, wie auch zu T1 und T2, für den Bereich Quellenverwechslungen nachweisen. Wiederum sind die Versuchspersonen der RM – Gruppe auch bei einer ersten Befragung zur Dia-Serie 'PKW-PKW' nach acht Wochen sehr gut bzw. besser als die der ESM – Gruppe in der Lage zwischen den Serien und Quellen zu differenzieren, was sich in einem signifikanten Ergebnis zeigt. Insgesamt und innerhalb der Serien ergaben sich wieder keine statistisch bedeutsamen Unterschiede zwischen den Versuchsgruppen.

Im statistischen Ergebnis gleichen die folgenden Befunde für die Serien 'PKW-Fahrrad' zu NT2 (nach acht Wochen) exakt den Befunden zu T1, also ebenfalls einer ersten Befragung – nach einer Woche. Es ist lediglich in der Richtungstendenz der schematypischen Intrusionen ein Unterschied zu verzeichnen.

Bezüglich korrekt berichteter Details sowie Quellenverwechslungen waren keinerlei signifikante Unterschiede zwischen den Gruppen zu verzeichnen.

Die Jugendlichen, die sich die zweite Unfall-Serie 'PKW-Fahrrad' selbst ausdenken sollten, begingen hierbei im Interview immens mehr Fehler insgesamt sowie im Bereich allgemeiner Erinnerungsfehler und konfabulierten auch deutlich mehr als diejenigen Jugendlichen, die eine Dia-Serie 'PKW-Fahrrad' erinnern sollten.

Auch bei einer ersten Befragung nach acht Wochen (NT2) zu den Serien 'PKW-Fahrrad' ergab sich hinsichtlich des Vorkommens schematypischer Intrusionen ein deutlicher Unterschied zwischen den Versuchsgruppen, allerdings mit dem Unterschied dass im Gegensatz zu T1 und T2 nun die RM – Gruppe mehr Intrusionen vorbringt als die ESM – Gruppe.

Betrachtet man zusammenfassend die Unterschiede zwischen den Versuchsgruppen in der Gedächtnisleistung über die verschiedenen Befragungszeitpunkte und Dia- bzw. Unfall-Serien hinweg, ergibt sich ein auf der einen Seite teilweise zwar ein recht uneinheitliches Bild, auf der anderen Seite lassen sich doch eventuell einige Charakteristika herausstellen. Einige davon decken sich mit bisherigen Befunden, während andere konträre Ergebnisse dazu darstellen. Wie bereits dargestellt, hängt es von dem Ausmaß der Ähnlichkeit von verschiedenen Quellen ab, wie gut zwischen diesen Quellen differenziert werden kann. Weiterhin konnten Raye und Johnson (1980) anhand von zwei Experimenten zeigen, dass die Unterscheidung zwischen einer externalen und einer internalen Quelle einfacher ist als die Unterscheidung zwischen zwei externalen Quellen.

Betrachtet man im Rahmen der Analyse der korrekt berichteten Details die Gedächtnisleistungen der beiden Gruppen, so lassen sich mit Ausnahme des ersten Befragungszeitpunktes auf Nachfrage zu keinem Zeitpunkt bedeutsame Unterschiede feststellen derart, dass es für eine der beiden Gruppen leichter wäre zwischen zwei externalen Quellen bzw. einer externalen und einer internalen Quelle zu unterscheiden. Ein Einfluss einer Verzögerung der Befragung auf die korrekt erinnerten Details ist offensichtlich nicht so sehr von Bedeutung, die diesbezügliche Gedächtnisleistung erweist sich als relativ stabil über die Zeit. Dies gilt sogar für diejenigen Versuchspersonen, die nach acht Wochen das erste Mal befragt wurden.

Etwas anders verhalten sich die Befunde bei getrennter Betrachtung der einzelnen Dia- bzw. Unfall-Serien. Bezüglich der Dia-Serie 'PKW-PKW'. Hier ergibt sich ein eindeutiger Befund dergestalt, dass diejenigen Jugendlichen, die zwischen einer externalen Quelle und einer internalen Quelle (selbst ausgedachter Unfall) differenzieren müssen, bei Abfrage der einzelnen Dia-Serie 'PKW-PKW' über bzw. zu allen drei Zeitpunkten sowohl insgesamt wie auch auf Nachfrage wesentlich mehr korrekte Details berichten als diejenigen, die zwischen zwei externalen Quellen unterscheiden mussten. Die Jugendlichen der Bedingung 'reality monitoring' sind demnach über die Zeit konstant besser in der Lage, Details aus der Dia-Serie 'PKW-PKW' korrekt zu erinnern als diejenigen der Bedingung externes 'source monitoring'.

Wiederum anders gestalten sich die Ergebnisse hinsichtlich der Dia-Serie bzw. selbst erfundenen Unfall-Serie 'PKW-Fahrrad'. Hier gleichen sich die Befunde der Befragungszeitpunkte T1 und NT2, also wenn die Versuchspersonen das erste Mal befragt wurden – nach einer bzw. acht Wochen. Hier fanden sich keinerlei Gruppenunterschiede bezüglich insgesamt, im freien Bericht oder auf Nachfrage berichteter Einzelheiten. Bemerkenswert ist jedoch das Ergebnis hinsichtlich eines zweiten Befragungszeitpunktes T2 nach acht Wochen, also dann, wenn ein Teil der Versuchspersonen ein zweites Mal befragt wurde. Dann erinnern bzw. berichten diejenigen Jugendlichen, die sich die Unfall-Serie 'PKW-Fahrrad' selbst ausdenken mussten und diese auch schriftlich niederlegten, deutlich weniger Details insgesamt sowie im freien Bericht als die Jugendlichen, die eine inhaltlich gleiche Dia-Serie sahen. Hier scheint also offensichtlich neben dem erwartbaren negativen Effekt der Zeit (Vergessen) auch eine zweite Befragung bzw. eben die Häufigkeit von Befragungen eine entscheidende Rolle zu spielen. Die Jugendlichen der Bedingung 'external-internal' sind somit deutlich schlechter in der Lage bei einer zweiten Befragung von sich aus diejenigen Einzelheiten zu erinnern, die sie sich acht Wochen zuvor selbst ausgedacht hatten, als die Jugendlichen der Bedingung 'external-external', die eine entsprechende Dia-Serie sahen. Bei einer ersten Befragung – nach einer, aber auch nach acht Wochen – zeigen sich diese Befunde nicht.

Hinsichtlich der Fehleranalyse bzw. der Art der begangenen Fehler über alle Befragungszeitpunkte ergeben sich auch ebenso klare wie auch weniger eindeutige Befunde.

Gut vergleichbar und eindeutig sind die Befunde der beiden Befragungszeitpunkte acht Wochen nach Präsentation (T2 und NT2) hinsichtlich der insgesamt begangenen Fehler sowie der allgemeinen Erinnerungsfehler. Diejenigen Jugendlichen der Versuchsgruppe 'external-internal', die also zwischen einer externalen und einer internalen (selbst ausgedachten) Quelle differenzieren mussten, begingen sowohl bei einer ersten wie auch bei einer zweiten Befragung nach acht Wochen jeweils deutlich mehr Fehler in den genannten Bereichen als die Jugendlichen der Vergleichsgruppe (externales source monitoring) und zwar sowohl bei Betrachtung über beide Serien hinweg als auch einzeln für die Serien 'PKW-PKW' und 'PKW-Fahrrad'. Bezüglich dieser Fehlerarten scheint also die zeitliche Verzögerung einer Befragung für Personen, die zwischen einer externalen und einer internalen Quelle unterscheiden müssen, wesentlich folgenschwerer zu sein, was sich in wesentlich fehlerbehafteteren Schilderungen niederschlägt.

Bei Betrachtung einer ersten von zwei Befragungen nach einer Woche lassen sich die eben dargestellten Befunde hinsichtlich der Gesamtfehlerzahl mit Ausnahme der Dia-Serie 'PKW-PKW' bestätigen. Auch bereits nach einer Woche unterlaufen den Jugendlichen der 'reality monitoring' - Gruppe wesentlich mehr Fehler insgesamt. Tendenziell, aber eben nicht statistisch bedeutsam, gilt dieser Befund auch für die Dia-Serie 'PKW-PKW'.

Hinsichtlich allgemeiner Erinnerungsfehler zeigen sich nach einer Woche lediglich für die Serie 'PKW-Fahrrad' Unterschiede zwischen den Versuchsgruppen. Offensichtlich erinnern Jugendliche, die sich diese Unfall-Serie selbst ausdenken mussten, bereits nach einer Woche wesentlich weniger derartige Einzelheiten (bzw. vergessen diese) im Gegensatz zu den Versuchspersonen, die die Dia-Serie sahen. Über beide Serien hinweg und die von beiden Versuchsgruppen gesehene Dia-Serie 'PKW-PKW' ergaben sich diesbezüglich nach einer Woche noch keine Unterschiede.

Bei Betrachtung der begangenen Quellenverwechslungen ergeben sich hingegen recht eindeutige Befunde:

Für die insgesamt sowie innerhalb der Versuchsmaterialien begangenen Quellenverwechslungen konnten zu keinem der Befragungszeitpunkte (T1, T2 oder NT2) - sei es über die beiden Serien hinweg oder auch bezüglich der einzelnen Dia- bzw. Unfall-Serie - Unterschiede festgestellt werden.

Bedeutsam und aussagekräftig erweisen sich zu allen Befragungszeitpunkten die Ergebnisse der begangenen Quellenverwechslungen zwischen den Serien. Bei Abfrage der Dia-Serie 'PKW-PKW' begingen diejenigen Versuchspersonen, die zwischen zwei externalen Quellen differenzieren sollten, wesentlich mehr Quellenverwechslungen zwischen den Dia-Serien.

Somit fiel denjenigen Jugendlichen, die nur eine Dia-Serie sahen und sich eine andere ausdachten, die Unterscheidung dieser beiden Quellen (external bzw. internal) sowohl über die Zeit hinweg konstant wie auch unabhängig von der Anzahl der Befragungen wesentlich leichter als denen der Vergleichsgruppe 'external-external', die jeweils deutlich mehr Quellenverwechslungen zwischen „ihren beiden Dia-Serien“ begingen.

Ebenso auffällig ist der Befund, dass sich bei Abfrage der Dia- bzw. Unfall-Serie 'PKW-Fahrrad' zu keinem der Befragungszeitpunkte wesentliche Unterschiede zwischen den Versuchsgruppen zeigten, was den Schluss zulässt, dass beide Versuchsgruppen sich der Elemente ihrer jeweiligen Serie recht bewusst waren und diese nicht mit Elementen der anderen Dia-Serie verwechselten.

Als besonders interessant erweisen sich die Befunde hinsichtlich begangener Konfabulationen und schematischer Intrusionen, da sich diese offensichtlich nach einem bestimmten Muster verhalten bzw. in einem wechselseitigen Verhältnis zueinander stehen. Dies gilt zumindest für die Befragungszeitpunkte T1 und T2, also für diejenigen Versuchspersonen, die zweimal interviewt wurden. Für einen ersten Befragungszeitpunkt NT2 nach allerdings erst acht Wochen scheinen besondere Bedingungen zu gelten:

Zunächst kann bezüglich des Vorkommens schematischer Intrusionen festgestellt werden, dass Jugendliche der Versuchsbedingung 'external-external' bei ihren beiden Befragungsterminen (T1 wie auch T2) über beide Dia-Serien hinweg wie auch bei Abfrage der Dia-Serie 'PKW-Fahrrad' deutlich mehr Intrusionen in ihre Schilderungen produzieren als diejenigen Jugendlichen der Bedingung 'external-internal'. Hinsichtlich der Dia-Serie 'PKW-PKW' finden sich zu beiden Zeitpunkten keine Unterschiede.

Genau umgekehrt gestalten sich offenbar die Befunde bzw. die Zusammenhänge bezüglich konfabulierter Inhalte. In exakt derselben Weise, also sowohl über beide Serien hinweg als auch hinsichtlich der Unfall-Serie 'PKW-Fahrrad', konfabulieren zu beiden Befragungszeitpunkten T1 und T2 allerdings die Jugendlichen der RM – Gruppe deutlich mehr als die Jugendlichen der ESM – Gruppe.

Bezüglich dieser beiden Fehlerarten kommt offenbar einem einmaligen ersten Befragungszeitpunkt NT2, der nach acht Wochen stattfand, eine interessante und spezielle Bedeutung zu. Hier kommen in den Berichten der 'reality monitoring' – Gruppe sowohl deutlich mehr schematische Intrusionen wie auch Konfabulationen vor als in den Berichten der externen 'source monitoring' – Gruppe hinsichtlich beider Serien simultan betrachtet wie auch der Unfall-Serie 'PKW-Fahrrad'. Eine kleine Ausnahme, d.h. kein signifikantes Ergebnis, bildet hier der Befund hinsichtlich konfabulierter Inhalte über beide Serien hinweg. Allerdings ist auch hier eine deutliche Tendenz beobachtbar, dass die Jugendlichen der Bedingung 'external-internal' häufiger konfabulieren als die Vergleichsgruppe.

Bezüglich der Dia-Serie 'PKW-PKW' konnten auch zu diesem Befragungszeitpunkt NT2 keine Unterschiede zwischen den beiden Versuchsgruppen im Bereich der Konfabulationen nachgewiesen werden.

5.1.2 Diskussion zu den Ergebnissen der Veränderung der Gedächtnisleistung über die Zeit

Im Folgenden werden die Veränderungen der Gedächtnis- und Erinnerungsleistung über die Zeit hinweg (T1 hin zu T2) betrachtet, zunächst für die externale 'source monitoring' - Gruppe, im Anschluss daran bezüglich der 'reality monitoring' – Gruppe.

Dabei wird analog zu obigen Ausführungen jeweils zunächst mit der Betrachtung über beide Dia- bzw. Unfall-Serien begonnen und daraufhin die Befunde der Dia-Serie 'PKW-PKW' und der Dia- bzw. Unfall-Serie 'PKW-Fahrrad' erläutert und diskutiert.

Betrachtet man die Ergebnisse der zeitlichen Veränderungen der Gedächtnisleistung getrennt nach den beiden Gruppen, so zeigt sich sowohl für die externale 'source monitoring' – Gruppe als auch für die 'reality monitoring' – Gruppe eine grobe bzw. nur in Teilbereichen gültige Replizierung bereits bekannter Befunde. Die Veränderung der Gedächtnisleistung vom ersten zum zweiten Befragungszeitpunkt folgt bei jeder der Gruppen zumindest tendenziell, aber auch in vielen wichtigen Bereichen (vor allem der Fehlerzahl) statistisch bedeutsam den gedächtnispsychologischen Gesetzmäßigkeiten dahingehend, dass die Versuchspersonen zum zweiten Befragungszeitpunkt insgesamt weniger korrekte Details und insgesamt mehr Fehler berichten. Es ergeben sich doch zum Teil deutliche Unterschiede in den beiden Versuchsgruppen, die im Einzelnen dargestellt und diskutiert werden.

externale 'source monitoring' - Gruppe:

Bezogen auf beide Dia-Serien ergaben sich hier keinerlei signifikante Veränderungen über die Zeit hinweg weder bezüglich der Gesamtzahl der korrekt berichteten Details oder im freien Bericht noch auf Nachfrage.

Auch hinsichtlich der Arten von Quellenverwechslungen (insgesamt, innerhalb und zwischen den Dia-Serien) ergeben sich über beide Dia-Serien hinweg keine wesentlichen Veränderungen zwischen den beiden Befragungszeitpunkten.

Gleiches gilt für schematypische Intrusionen sowie auch Konfabulationen. Zum zweiten Befragungszeitpunkt T2 unterlaufen den Versuchspersonen tendenziell betrachtet weniger Intrusionen, dafür allerdings etwas mehr Konfabulationen als zu T1, was statistisch betrachtet jedoch in beiden Fällen ohne Relevanz bleibt.

Dagegen werden für beide Dia-Serien zusammen zu T2 hochsignifikant mehr Fehler insgesamt begangen als zu T1. Analog dazu nimmt auch die Rate der allgemeinen Erinnerungsfehler von T1 nach T2 deutlich zu.

Trotz einiger leichter tendenzieller Bewegungen ließen sich bemerkenswerterweise für die *Dia-Serie 'PKW-PKW'* weder hinsichtlich der korrekt berichteten Details noch der Fehler und Fehlerarten wesentlich Veränderungen zwischen den Befragungszeitpunkten beobachten. Für keine der Variablen fand sich ein signifikantes Ergebnis.

Tendenziell war für folgende Variablen ein leichter Anstieg in der Ausprägung des jeweiligen Mittelwerts vom ersten zum zweiten Befragungszeitpunkt zu verzeichnen:

korrekte Details 'gesamt' und im 'free recall', Fehler 'gesamt', Quellenverwechslungen 'gesamt' und innerhalb der Dia-Serie, Konfabulationen und allgemeine Erinnerungsfehler. Dagegen verzeichneten 'korrekte Details auf Nachfrage', 'Quellenverwechslungen zwischen den Dia-Serien' sowie 'schematypische Intrusionen' dagegen tendenziell zum zweiten Befragungszeitpunkt eher geringere Ausprägungen als zum ersten.

Bezüglich der *Dia-Serie 'PKW-Fahrrad'* ähneln sich die Befunde gegenüber der Dia-Serie 'PKW-PKW' mit Ausnahme lediglich der Ergebnisse hinsichtlich der Gesamtfehlerzahl sowie der allgemeinen Erinnerungsfehler. So werden in dieser Versuchsgruppe zum zweiten Befragungszeitpunkt T2 sowohl deutlich mehr Fehler insgesamt als auch allgemeine Erinnerungsfehler begangen als zu T1.

Hinsichtlich der korrekt berichteten Details sowie der übrigen Fehlerarten finden sich in dieser zweiten Dia-Serie keine bedeutsamen Unterschiede zwischen erstem und zweitem Befragungszeitpunkt.

Lediglich leichte tendenzielle Veränderungen in Form eines Anstiegs bzw. Abfalls von T1 nach T2 konnten für korrekte Details gesamt und im freien Bericht, schematypische Intrusionen und Konfabulationen bzw. korrekte Details 'auf Nachfrage' beobachtet werden. Sämtliche Formen von Quellenverwechslungen zeigten auch keinerlei tendenzielle Veränderungen zwischen den Befragungszeitpunkten.

Neben einer Vielzahl feiner Tendenzen in verschiedene Richtungen weisen die Ergebnisse der Veränderung der Gedächtnisleistung der externalen 'source monitoring' - Gruppe über die Zeit, also von T1 nach T2 ein doch deutliches Befundmuster auf:

Im Bereich der korrekt berichteten Details ergeben sich für keine der Teilvariablen signifikante Veränderungen über die Zeit. Was die Anzahl der insgesamt, im freien Bericht oder auf Nachfrage erinnerten bzw. geschilderten Details angeht, scheint die Erinnerungsleistung hypothesenkonträr bzw. nicht unbedingt den üblichen Gesetzmäßigkeiten des Vergessens

entsprechend sowohl über beide Dia-Serien hinweg als auch hinsichtlich separater Analyse der einzelnen Dia-Serien über die Zeit hinweg recht konstant zu sein, also unabhängig vom zeitlichen Abstand der beiden Befragungszeitpunkte. Zum einen sei hier neben weiteren noch anzuführenden Aspekten als mögliche Erklärung das doch im Schnitt relativ hohe intellektuelle Niveau der Versuchspersonen genannt.

Hinsichtlich der durchgeführten Fehleranalyse kann jeweils analog zueinander festgestellt werden, dass mit fortschreitender Zeit sowohl die Fehlerrate der insgesamt begangenen Fehler wie auch der allgemeinen Erinnerungsfehler deutlich zunimmt. Dies gilt nicht für die Fehlerraten innerhalb der Dia-Serie 'PKW-PKW', obwohl dort tendenziell die Fehlerquoten zu T2 erwartungsgemäß auch höher sind.

Bezüglich aller übrigen Fehlerarten muss bzw. kann konstatiert werden, dass sich die Unterschiede zwischen den beiden Befragungszeitpunkten in einem relativ kleinen Rahmen bewegen, also recht konstant über die Zeit hinweg sind. Tendenziell weisen jedoch die meisten Befunde in die hypothesenkonforme Richtung einer Zunahme der jeweiligen Fehlerquote bzw. einer Verschlechterung oder größeren Störanfälligkeit der Erinnerungsleistung mit zunehmendem zeitlichen Abstand.

reality monitoring – Gruppe:

Im Gegensatz zur ESM – Gruppe berichten die Versuchspersonen der RM – Gruppe zum zweiten Befragungszeitpunkt T2 *über beide Unfall-Serien* hinweg deutlich weniger korrekte Details insgesamt und auch auf Nachfrage als zum ersten. Zu T2 werden im freien Bericht zwar etwas mehr korrekte Details berichtet als zu T1, jedoch ist dies statistisch unbedeutend. Wie auch die Jugendlichen der Versuchsgruppe 'external-external' machen die Jugendlichen der Gruppe 'external-internal' zum zweiten Befragungszeitpunkt T2 nicht nur deutlich mehr Fehler insgesamt, sondern auch deutlich mehr allgemeine Erinnerungsfehler als zum ersten Befragungszeitpunkt T1.

Die Veränderungen zwischen den Zeitpunkten bezüglich der Quellenverwechslungen erweisen sich auch für diese Versuchsbedingung erneut als dermaßen gering, dass kein statistisch bedeutsames Ergebnis festgehalten werden kann.

Ebensolches zeigt sich für das Auftreten schematypischer Intrusionen zu den beiden Zeitpunkten. Es ist lediglich ein leichter tendenzieller Abfall von T1 nach T2 erkennbar.

Dahingegen treten in der Gruppe 'reality monitoring' Konfabulationen jedoch zum zweiten Befragungszeitpunkt wesentlich häufiger auf als zum ersten.

Bei separater Betrachtung der *Dia-Serie 'PKW-PKW'* ergaben sich für die korrekt berichteten Details keine wesentlichen Veränderungen über die Zeit hinweg, obwohl sich tendenziell hier insgesamt und im freien Bericht zu T2 höhere Werte als zu T1 feststellen ließen. Die

Jugendlichen, die also insgesamt zwischen einer externalen und einer internalen Quellen differenzieren mussten, erinnerten zu einem späteren, zweiten Befragungszeitpunkt entgegen der erwartungsgemäßen Annahme mehr Details korrekt als zum ersten.

Im Gegensatz zur externalen 'source-monitoring'-Gruppe fanden sich hier für die Dia-Serie 'PKW-PKW' wiederum erwartungsgemäß zum zweiten Befragungszeitpunkt T2 sowohl wesentlich höhere Gesamtfehlerzahlen als auch wesentlich mehr allgemeine Erinnerungsfehler im Vergleich zu T1.

Was die Arten von Quellenverwechslungen aber auch vorgebrachte konfabulierte Inhalte betrifft, konnten wiederum kaum bzw. keine signifikanten Veränderungen zwischen den Befragungszeitpunkten festgestellt werden.

Zwar zeigt sich hinsichtlich schematischer Intrusionen ein signifikantes Ergebnis, jedoch kommen zu T2 weniger Intrusionen vor als zu T1.

Im Gegensatz zur Dia-Serie 'PKW-Fahrrad' der ESM – Gruppe berichten die Jugendlichen der RM – Gruppe zum zweiten Befragungszeitpunkt deutlich weniger korrekte Details (gesamt und auf Nachfrage) ihrer *selbst erfundenen Unfall-Serie 'PKW-Fahrrad'* als zum ersten Befragungszeitpunkt. Im freien Bericht sind dies tendenziell auch weniger, jedoch statistisch nicht bedeutsam.

Zur Dia-Serie 'PKW-Fahrrad' der ESM – Gruppe analoge Ergebnisse ließen sich jedoch hinsichtlich der Gesamtfehlerzahl sowie der Anzahl der allgemeinen Erinnerungsfehler in der selbst erfundenen Unfall-Serie 'PKW-Fahrrad' feststellen. In beiden Fällen stieg die Fehlerzahl von T1 zu T2 sehr deutlich an.

Erneut veränderten sich Quellenverwechslungen und auch schematische Intrusionen bei Betrachtung der selbst erfundenen Unfall-Serie 'PKW-Fahrrad' über die Zeit fast nicht und blieben im statistischen Ergebnis ohne Bedeutung.

Allerdings konfabulierten die Jugendlichen zu T2 signifikant mehr als zu T1, sollten sie sich eine der Dia-Serie 'PKW-Fahrrad' analoge Unfall-Serie selbst ausdenken.

Bei Betrachtung der Gedächtnisleistung der 'reality monitoring' - Gruppe über die Zeit hinweg werden die meisten Ergebnisse der entsprechenden Befunde der externalen 'source monitoring'-Gruppe nicht nur bestätigt, sondern auch um bzw. in einigen Aspekten ergänzt: Während die Gedächtnisleistung über die Zeit hinsichtlich des Erinnerns bzw. Berichtens korrekter Details bei der 'externalen source monitoring' - Gruppe noch relativ stabil ist, sinkt erwartungsgemäß und den üblichen Gedächtnisprozessen und gedächtnispsychologischen Gesetzmäßigkeiten (Vergessen über die Zeit) entsprechend die Anzahl der zum zweiten Befragungszeitpunkt T2 - insgesamt wie auch auf Nachfrage hin - korrekt berichteten Details. Dies gilt jedoch nur bei Betrachtung über beide Serien hinweg und für die selbst erfundene

Unfall-Serie 'PKW-Fahrrad'. Offensichtlich erinnern Jugendliche, die zwischen einer externalen und einer internalen Quelle differenzieren müssen, die selbst generierte Quelle zum einen über die Zeit hinweg schlechter als auch im Vergleich zu der externalen Quelle schlechter. Hinsichtlich der Dia-Serie 'PKW-PKW' ergaben sich zwar keinerlei statistisch bedeutsame Veränderungen über die Zeit hinweg, jedoch sei auf ein interessantes Ergebnis hingewiesen. In dieser Versuchsgruppe steigt tendenziell die Anzahl der insgesamt sowie auch im freien Bericht korrekt geschilderten Details vom ersten Befragungszeitpunkt zu einem zweiten hin an. Möglicherweise erinnern die Jugendlichen ihre selbst erfundene Unfall-Serie so schlecht, dass sie sich sowohl in der Enkodierungsphase wie auch beim Abruf der gesehenen Dia-Serie verstärkt auf diese konzentrieren und im Sinne von Reminiscenz- und Hypermneseeffekten auch bei einer zweiten Befragung von allgemeinen Lerneffekten und Kenntnis der Testsituation (*test sophistication*) von dieser Befragung profitieren. Die Versuchspersonen, unabhängig aus welcher Gruppe, konnten aus den Erfahrungen des ersten Interviews lernen, das heißt, sie kannten zum zweiten Testzeitpunkt bereits den Ablauf der Befragungen und konnten so unbefangener in das Interview hineingehen als noch zum ersten Befragungszeitpunkt. Man könnte also vermuten, dass die Gewöhnung an die Befragungssituation dazu geführt hat, dass die Versuchspersonen - zumindest der 'reality monitoring' - Gruppe - beim zweiten Interview nicht mehr so nervös waren und die anfängliche Unsicherheit verringert werden konnte, was sich beispielsweise darin bemerkbar machte, dass teilweise die freien Berichte bedeutend länger ausfielen als noch beim ersten Interview. Schon während der Instruktion und Stimuluspräsentation wurde jedoch deutlich, dass viele der Jugendlichen schon Schwierigkeiten hatten, sich trotz klarer Vorgaben eine entsprechende Unfall-Serie auszudenken und dann später zu erinnern.

Des Weiteren könnte eine Verbesserung der Gedächtnisleistung darauf zurückzuführen sein, dass die Versuchspersonen nach der ersten Befragung acht Wochen Zeit hatten, sich gedanklich auf den zweiten Testzeitpunkt vorzubereiten. Dabei könnte es auch zu einem Austausch zwischen den Versuchspersonen gekommen sein, der dazu führte, dass einige Versuchspersonen beim zweiten Befragungszeitpunkt ihre Angaben aus dem ersten Interview revidierten und so mehr korrekte Details berichteten. Diese Möglichkeit bestand zwar und für die meisten der Jugendlichen waren die Befragungssituationen auch eine teils spannende Erfahrung, aber auf jeden Fall eine willkommene Abwechslung zu ihrem Alltag. Jedoch ist auch klar, dass gerade bei Jugendlichen, die nicht allzu sehr intelligenzgemindert sind, eine motivationale Komponente entscheidend ist. Da es sich um recht neutrales, d.h. nicht unbedingt persönlich-emotional erschütterndes Versuchsmaterial handelte, dürfte sich der Stellenwert solcher Befragungen für die Jugendlichen aber in einem geringen Rahmen bewegen, so dass ernsthafte, sich deutlich auswirkende Absprachen oder ein Austausch von

Informationen bezüglich der Dia-Serien eher auszuschließen sind. Zudem wussten die meisten der Jugendlichen nicht genau, wann die nächste Befragung stattfindet. Somit sind nach Meinung des Verfassers die meisten Arten von suggerierten Falschinformationseffekten in der vorliegenden Arbeit eher weniger als mögliche Erklärungen relevant. Demgegenüber stehen die Ergebnisse der Untersuchung von HENRY & GUDJONSSON (2003), wonach es bei 11 und 12-Jährigen mit und ohne Lernbehinderung zum Anstieg der freien Erinnerungsleistung über zwei Interviews hinweg kam, jedoch nicht hinsichtlich spezifischer geschlossener Fragen. Im Fall einer gleich bleibenden oder über die Zeit abnehmenden Erinnerungsleistung folgt daraus nicht notwendigerweise, dass die gleiche Information einfach von Interview zu Interview übernommen wird.

Wiederum eindeutig sind die Befunde der Veränderung über die Zeit hinsichtlich der insgesamt begangenen Fehler wie auch der allgemeinen Erinnerungsfehler. Konnte allerdings in der Gruppe 'external-external' keine Veränderung von T1 nach T2 in Bezug auf die Dia-Serie 'PKW-PKW' beobachtet werden, kommen in den Berichten der Jugendlichen der Versuchsbedingung 'external-internal' für beide Serien zusammengekommen wie auch einzeln betrachtet zum zweiten Befragungszeitpunkt T2 jeweils wesentlich höhere Fehlerquoten der genannten Fehlerarten vor. Diese zeitlichen Veränderungen in den Fehlerarten erweisen sich für die Versuchsgruppe 'reality-monitoring' allerdings als noch größer als für die Vergleichsgruppe.

Exakt identisch mit den Befunden der Versuchsgruppe externes 'source monitoring' gestalten sich auch die Befunde der 'reality monitoring' - Gruppe hinsichtlich der Veränderung begangener Quellenverwechslungen (sei es insgesamt, innerhalb oder zwischen den Serien) vom ersten zum zweiten Befragungszeitpunkt. Auch hier finden sich keinerlei wesentlichen Unterschiede zu T1 im Vergleich zu T2. Die insgesamt sehr geringe Rate an Quellenverwechslungen scheint offensichtlich auch relativ unabhängig vom zeitlichen Verlauf zu sein. Demgegenüber stehen die Ergebnisse von HENKEL (2004), die fand, dass Versuchspersonen über mehrere Erinnerungsversuche hinweg mehr Fehler in Form von Quellenverwechslungen unterliefen bei der Entscheidung, ob die zu erinnernden Stimuli als Bilder oder Worte präsentiert wurden. Dies galt insbesondere dann, wenn die Versuchspersonen sich nicht bewusst waren, dass sie später zu der Quelle der Erinnerung befragt werden. Versuchspersonen waren allerdings nicht lernbehinderte Personen. In der Untersuchung von HENKEL (2004) zeigten sich neben einer Zunahme der erinnerten Items mit wiederholten Befragungen ebenso vermehrt Quellenfehler. Dabei decken sich die genannten Befunde einerseits auch mit den Befunden der vorliegenden Arbeit sowie mit der allgemeinen Einschätzung der möglichen gegensätzlichen Effekte wiederholter Befragungen auf die Ge-

dächtnisleistung (*'paradoxical effects of repeated testing'*). Andererseits stehen sie auch im Widerspruch dazu. Ein Anstieg an erinnerten Details geht demnach immer mit vermehrten Quellenverwechslungen dieser erinnerten Items einher. Bei Erwachsenen waren nach Bornstein et al. (1998; zit. nach HENKEL, 2004) wiederholte Erinnerungsversuche von Einzelheiten eines Verbrechens zwar sowohl mit dem Erinnern mehr korrekter Informationen verbunden wie jedoch auch mit einer (geringen, aber doch deutlichen) Zunahme von Intrusionen. Solche Erinnerungsfehler könnten systematische Verzerrungen zum Ausdruck bringen, die Teil normaler kognitiver Prozesse sind. Somit sind diesbezüglich immer die Kosten gegenüber dem Nutzen wiederholter Befragungen in Form der Fehleranzahl gegenüber der Zahl der korrekten Details gegeneinander abzuwägen.

Während sich beim Vergleich über die Befragungszeitpunkte hinsichtlich schematypischer Intrusionen und Konfabulationen für die Jugendlichen der Bedingung 'external-external' keinerlei Veränderungen für keine der beiden Dia-Serien ergaben, verhält sich dies für die Versuchspersonen der anderen Bedingung 'external-internal' etwas differenzierter. Fehler in Form von schematypischen Intrusionen finden sich zu einem zweiten Befragungszeitpunkt lediglich für die gesehene Dia-Serie 'PKW-PKW' weniger als zu einem ersten, dagegen nicht in der Schilderung der selbst erfundenen Unfall-Serie 'PKW-Fahrrad'. Trotz des offensichtlich erschwerten Zugangs oder Abruf der intern generierten Quelle, finden sich dennoch zu T2 nicht mehr solcher Fehler als zu T1 oder umgekehrt. Damit in Zusammenhang stehen nach Meinung von THOMSON & POWELL (2003) die bei Kindern und intellektuell beeinträchtigten Personen kritisch zu betrachtende Fähigkeit, zwischen ähnlichen Vorfällen und spezifischen Details wiederholter Ereignisse differenzieren zu können, insbesondere nach einer zeitlichen Verzögerung. Gerade bei diesen Personengruppen stellt dies ein komplexes Problem dar, da deren Gedächtnis- und Erinnerungsprozesse nicht so gut entwickelt sind wie die von Erwachsenen. Fehler treten hier in erster Linie in Form von internalen Intrusionen, also der Übertragung bzw. Integration von Details anderer Vorfälle auf bzw. in den aktuell zu Erinnernden Vorfall. Je jünger die Kinder und je länger das Behaltensintervall, desto größer die Anzahl internaler Intrusionsfehler. Berichtete externale Intrusionen (falsche Details die in keinem der Vorfälle des Ereignisses vorkamen) sind von Kindern, die ein Ereignis wiederholt erfahren haben, weniger zu erwarten. Obwohl Alter und Behaltensintervall ebenfalls die Fähigkeit von Kindern beeinflusst ein einzelnes Ereignis zu erinnern, ist der Einfluss der genannten Faktoren größer, wenn Kinder Vorfälle eines wiederholten, veränderlichen Geschehens erinnern sollen. Sowohl bei der Schilderung einzelner wie auch wiederholter Ereignisse legen die Befunde der Forschung nahe, dass eine Spezifizierung der Fragestellung einerseits mit einem Anstieg korrekter wie auch falscher Details einhergeht. Die Fehlerzahl steigt deshalb, da bei einer Abfrage im Falle nicht vorhandener oder nicht

zugänglicher Informationen, das Kind raten muss bzw. eine Antwort generiert, die im Interesse des Befragenden liegt, dessen Vorstellungen entspricht oder einfach eine auf ähnliche Erfahrungen bezogene Antwort gibt, um nicht keine Antwort zu geben. Im freien Bericht ergeben sich dabei auch immer mehr Anhaltspunkte, die die einzelnen Vorfälle innerhalb der Ereignisse voneinander unterscheidbar machen. Da bei lernbehinderten Jugendlichen oder generell kognitiv eingeschränkten Personen egal welchen Alters kaum empirische Befunde diesbezüglich vorliegen, besteht auch hier dringend Forschungsbedarf, auch auf dem Gebiet der Grundlagenforschung, die darüber hinaus verstärkt auch motivationale Aspekte beachten sollte.

Demgegenüber finden sich in der vorliegenden Untersuchung deutlich mehr konfabulierte Inhalte zu T2 als zu T1 bei Betrachtung über beide Unfall-Serien zusammen wie auch vor allem für die intern generierte Quelle (selbst erfundene Unfall-Serie 'PKW-Fahrrad'). Offensichtlich wird der mangelnde Zugang bzw. Abruf der intern generierten Quelle zu einem zweiten Befragungszeitpunkt T2 mehr als zu einem ersten versucht mit Konfabulationen zu überspielen bzw. fehlende Inhalte aufzufüllen.

5.1.3 Diskussion zu den Ergebnissen der Gedächtnisleistungen zu den einzelnen Befragungszeitpunkten

Im Folgenden werden die Befunde der beiden Versuchsgruppen getrennt voneinander im Vergleich der beiden jeweiligen Befragungszeitpunkte (T1-NT2 und T2-NT2) hin betrachtet und diskutiert.

Begonnen wird mit der Betrachtung der Zeitpunkte T1-NT2, gefolgt vom Vergleich T2-NT2. Dargestellt werden jeweils zuerst die Befunde der Gruppe externes 'source monitoring', daraufhin die der Gruppe 'reality monitoring'.

Dabei werden analog zu obigen Ausführungen zunächst jeweils die Ergebnisse über beide Dia- bzw. Unfall-Serien hinweg erläutert sowie jeweils getrennt voneinander ('PKW-PKW' und 'PKW-Fahrrad').

Vergleich T1-NT2: Gruppe 'source monitoring':

Beim Vergleich der korrekten Details berichten die Jugendlichen bezogen auf *beide Dia-Serien* insgesamt wie auch im 'free recall' nach einer Woche zu T1 deutlich mehr Details als nach acht Wochen zu NT2. Auf Nachfrage sind es tendenziell zwar auch mehr, jedoch ist dieser Unterschied statistisch nicht-signifikant.

Wie zu vermuten war, ist die Gesamtfehlerzahl sowie die Rate der allgemeinen Erinnerungsfehler bei Befragung nach acht Wochen (NT2) zwar höher als nach einer Woche (T1),

allerdings erweisen sich beide Unterschiede statistisch gesehen als unbedeutend – bezogen auf beide Dia-Serien.

Ebenso nicht-signifikant zeigt sich ein weiteres Mal der Vergleich der begangenen Quellenverwechslungen. Auch hier lassen sich über beide Dia-Serien hinweg keine wesentlichen Veränderungen zwischen T1 und NT2 nachweisen.

Hochsignifikante Veränderungen ergaben sich jedoch sowohl hinsichtlich schematypischer Intrusionen als auch bezüglich Konfabulationen - allerdings in unterschiedliche Richtungen:

Während die Versuchspersonen bei Abfrage zu T1 in beiden präsentierten Dia-Serien mehr schematypische Intrusionen vorbrachten als zu NT2, wurde zu NT2 deutlich mehr konfabuliert als zu T1.

Wie schon beim Vergleich der Gedächtnisleistungen über die Zeit (T1-T2) ergibt sich bei isolierter Betrachtung der *Dia-Serie 'PKW-PKW'* für die ESM – Gruppe hier ebenfalls für keine der Variablen eine signifikante Veränderung zwischen den Befragungszeitpunkten T1 und NT2.

Es lassen sich hier auch nur eventuelle Richtungstendenzen beobachten. Zu NT2 ergaben sich für die Gesamtfehlerzahl, Quellenverwechslungen innerhalb der Dia-Serie, Konfabulationen sowie allgemeine Erinnerungsfehler höhere Werte als zu T1, umgekehrt gilt dies für korrekte Details 'gesamt', im 'free recall' und 'auf Nachfrage', Quellenverwechslungen insgesamt und zwischen den Serien sowie schematypische Intrusionen.

Völlig anders als beim Vergleich T1-T2 über die Zeit lassen sich bei Betrachtung der *Dia-Serie 'PKW-Fahrrad'* für sich allein für die Versuchspersonen der Bedingung 'external-external' sehr deutliche Unterschiede zwischen den Befragungszeitpunkten T1 und NT2 aufzeigen. Hier berichten die Jugendlichen insgesamt, im 'free recall' und auch auf Nachfrage bei einer ersten Befragung nach acht Wochen jeweils deutlich weniger korrekte Details als bei einer ersten Befragung nach einer Woche.

Im Gegensatz zum Vergleich T1-T2 der ESM – Gruppe diesbezüglich gestalten sich die Befunde hinsichtlich der Gesamtfehlerzahl sowie der Zahl der allgemeinen Erinnerungsfehler. Beide Fehlerraten unterscheiden sich zu den Befragungszeitpunkten T1 und NT2 kaum und nicht wesentlich voneinander. Gleiches gilt innerhalb der Dia-Serie 'PKW-Fahrrad' für das Vorkommen von Quellenverwechslungen und Konfabulationen.

Ein signifikanter Unterschied zwischen T1 und NT2 findet sich lediglich innerhalb der schematypischen Intrusionen; jedoch ist hier zu bemerken, dass diese für die Dia-Serie 'PKW-Fahrrad' nach acht Wochen weniger häufig vorkommen als nach einer Woche.

Vergleich T1-NT2: Gruppe 'reality monitoring':

Anders als beim Vergleich T1-NT2 für die ESM – Gruppe berichten die Versuchspersonen der RM - Gruppe über beide Unfall-Serien hinweg – zwar tendenziell zu NT2 ebenfalls weniger als zu T1 - signifikant allerdings nicht weniger korrekte Details insgesamt, im freien Bericht und auch auf Nachfrage hin.

Wiederum erwartungsgemäß ergibt sich das Bild für die insgesamt begangenen sowie die allgemeinen Erinnerungsfehler. Auch die RM – Gruppe macht bei einer ersten Befragung nach acht Wochen wesentlich mehr dieser Fehler als bei einer ersten Befragung, die schon nach einer Woche stattfindet – bezogen auf beide Unfall-Serien. Gleiches gilt auch erneut für die Fehlerart 'schematypische Intrusionen'.

Kaum feststellbare Unterschiede zwischen T1 und NT2 sind bezüglich der Quellenverwechslung insgesamt, innerhalb und zwischen den Serien sowie konfabulierten Inhalten beider Unfall-Serien zu bemerken.

Bei wiederum isolierter Betrachtung der *Dia-Serie 'PKW-PKW'* macht es für die Jugendlichen der Versuchsbedingung 'external-internal' offensichtlich keinen Unterschied hinsichtlich des Erinnerns korrekter Details, ob eine erste Befragung bereits nach einer Woche (T1) oder erst nach ca. acht Wochen (NT2) stattfindet. Hier finden sich keinerlei wesentliche Unterschiede, sei es hinsichtlich der Gesamtzahl der erinnerten Details oder im freien Bericht oder auf Nachfrage.

Dem gegenüber scheinen auch hier wieder die Fehlerraten entscheidend zu sein. Sowohl in der Gesamtfehlerzahl als auch was allgemeine Erinnerungsfehler angeht werden in der Dia-Serie 'PKW-PKW' zu NT2 deutlich und höchstsignifikant mehr Fehler begangen als zu T1.

Keinerlei signifikante Unterschiede sind für die Dia-Serie 'PKW-PKW' hinsichtlich aller Formen von Quellenverwechslungen sowie Konfabulationen zu konstatieren. Tendenziell sind die diesbezüglichen Fehlerwerte erwartungsgemäß zu NT2 jedoch höher als zu T1.

Schematypische Intrusionen kommen für diese Dia-Serie allerdings nach acht Wochen signifikant häufiger vor als nach einer Woche.

Betrachtet man dagegen die *selbst erfundene Unfall-Serie 'PKW-Fahrrad'* per-se, kann hinsichtlich der korrekt berichteten Details in allen Abstufungen (gesamt, im freien Bericht und auf Nachfrage) festgestellt werden, dass diejenigen Jugendlichen, die nach einer Woche erstmals befragt werden deutlich mehr Details korrekt erinnern als diejenigen, die nach ca. acht Wochen erstmals befragt wurden.

Analog dazu unterlaufen den zuletzt genannten auch wesentlich mehr Fehler insgesamt wie auch im Bereich der allgemeinen Erinnerungsfehler.

Bezüglich der restlichen Fehlerarten (Quellenverwechslungen, Intrusionen und Konfabulationen) lassen sich nur tendenzielle Veränderungen in Form von zu NT2 höheren Werten als zu T1 beobachten. Für ein statistisch relevantes Ergebnis sind diese Unterschiede zu gering oder beruhen lediglich auf zufälligen Effekten.

Hinsichtlich der Frage, ob sich Unterschiede in der Gedächtnisleistung ergeben, wenn eine erste Befragung nach einer (T1) bzw. erst nach acht Wochen (NT2) stattfindet, zeigen sich für beide untersuchten Gruppen bei paralleler Betrachtung interessante, teilweise miteinander vereinbare oder vergleichbare, teilweise aber auch völlig unterschiedliche Ergebnisse:

Bei Analyse der korrekt berichteten Details, fällt zunächst auf, dass sich die Erinnerungsleistung beider Versuchsgruppen hinsichtlich der gesehenen Dia-Serie 'PKW-PKW' zu den genannten Zeitpunkten nicht wesentlich unterscheiden. Es macht also beim Abruf der externalen Quelle 'PKW-PKW' offensichtlich weder für die Jugendlichen, die zwischen zwei externalen Quellen differenzieren mussten wie auch für die Jugendlichen, die zwischen einer intern generierten und einer externalen Quelle unterscheiden sollten, keinen großen Unterschied, ob diese erste Abfrage bereits nach einer Woche oder erst nach acht Wochen stattfindet. Der Abruf und Zugang dieser bildhaften, gesehenen Repräsentation (Dia-Serie) ist offensichtlich recht stabil und auch für offene wie geschlossene Fragen konstant.

Ebenso eindeutig fallen die diesbezüglichen (also auf das Erinnern bzw. Berichten korrekter Details) Befunde für beide Versuchsgruppen hinsichtlich der Dia- bzw. Unfall-Serie 'PKW-Fahrrad' aus. Der Zeitpunkt einer ersten Befragung spielt hier eine entscheidende Rolle: Bei einer ersten Befragung nach einer Woche (T1) berichten im Vergleich zu einer ersten Befragung nach acht Wochen (NT2) die Jugendlichen, die eine entsprechende Dia-Serie gesehen hatten (external) ebenso wie die Jugendlichen, die sich eine entsprechende Unfall-Serie ausgedacht hatten (internal) wesentlich mehr korrekte Details insgesamt, im freien Bericht und auch auf Nachfrage. Somit zeigen sich hier beim Vergleich dieser beiden Zeitpunkte bedeutsame Unterschiede zwischen den Quellen einerseits und dem Einfluss des Zeitpunktes einer ersten Befragung.

Während sich in der Versuchsbedingung 'external- internal' über beide Serien zusammen betrachtet, keine Unterschiede zwischen T1 und NT2 zeigen, sind diese doch deutlich in der Versuchsbedingung 'external-external' zu finden, dergestalt, dass sich auch hier erwartungsgemäß – zumindest insgesamt wie auch im freien Bericht nach einer Woche deutlich mehr korrekt erinnerte Details finden als nach acht Wochen. Dabei ist dieser Befund vor allem auch auf die sehr deutlichen diesbezüglichen Unterschiede der Dia-Serie 'PKW-Fahrrad' zurückzuführen.

Wie in einigen vorangegangenen Vergleichen erweisen sich erneut die Gesamtfehlerzahl wie auch die Zahl der allgemeinen Erinnerungsfehler als entscheidende Kriterien. Vergleicht man nun eine erste Befragung nach einer Woche mit einer ebenfalls ersten Befragung nach acht Wochen, so lassen sich für die Versuchsgruppe 'external-external' weder über die beiden Dia-Serien hinweg noch für die beiden Dia-Serien einzeln wesentliche Veränderungen in den beiden Fehlerraten zwischen den Zeitpunkten feststellen, obwohl tendenziell zu NT2 jeweils mehr Fehler begangen wurden. Demgegenüber finden sich analog dazu in der Versuchsgruppe 'external-internal' bei einer ersten Befragung nach acht Wochen (NT2) jeweils deutlich höhere, statistisch relevante, Fehlerquoten. Somit ist es diesbezüglich für Jugendliche, die zwischen einer externalen und einer internalen Quelle unterscheiden müssen, sehr wohl von Bedeutung, wann eine erste Befragung stattfindet. Dies spricht in jedem Fall dafür in solchen Situationen eine Erstbefragung möglichst frühzeitig durchzuführen, um diese dann doch sehr hohen Fehlerquoten nach einigen Wochen zumindest geringer zu halten.

Ein weiteres Mal konnten auch beim Vergleich der Befragungszeitpunkte T1 und NT2 miteinander für keine der beiden Versuchsgruppen wesentliche Veränderungen hinsichtlich Quellenverwechslungen innerhalb oder zwischen den Dia- bzw. Unfall-Serien festgestellt werden. Somit scheinen Quellenverwechslungen nach einer wie nach acht Wochen im selben (geringen) Ausmaß vorzukommen bzw. eben nicht. Dieser Befund deckt sich mit den Ergebnissen einer Studie von Jens et al. (1990, zit. nach LORSBACH, 2000), wo geistig behinderte Kinder hinsichtlich ihrer Erinnerungs- und Diskriminationsleistung von selbst ausgeführten und nur vorgestellten Aktivitäten befragt wurden. Die Hälfte der Kinder wurde auch ein zweites Mal nach acht Wochen dazu befragt. Im Ergebnis fanden sich keine Gruppenunterschiede hinsichtlich Defizite bei Quellenidentifikationen, wobei beide Gruppen ähnliche Antwortmuster zeigten. Dabei erweis sich die Quellenzuordnung für selbst ausgeführte Aktivitäten im Allgemeinen als genauer als für zuvor nur vorgestellte.

Dagegen fand LORSBACH in seinen Untersuchungen übereinstimmende Befunde derart, dass Kinder mit Lernbehinderungen Schwierigkeiten haben, die Quellen ihrer Erinnerungen zu unterscheiden (LORSBACH & EWING, 1995; LORSBACH et al., 1998), da sie möglicherweise Probleme in der reflektiven Verarbeitung und bei verbalen Aufgaben haben (LORSBACH, 2000).

Unterschiedliche, komplexere Befunde lassen sich jedoch bei Betrachtung des Vorkommens schematischer Intrusionen wie auch Konfabulationen beobachten:

Während es bei denjenigen Jugendlichen, die zwischen zwei externalen Quellen oder Dia-Serien trennen mussten, sowohl über beide Dia-Serien hinweg wie auch hinsichtlich der Dia-Serie 'PKW-Fahrrad' zu deutlich mehr Intrusionen bei einer ersten Befragung nach einer Woche gegenüber nach acht Wochen kommt, verhält es sich in der Gruppe, in der zwischen

einer externalen und einer internalen Quelle unterschieden werden musste, so, dass ebenfalls über beide Serien hinweg wie jedoch auch für die Dia-Serie 'PKW-PKW' umgekehrt nach acht Wochen wesentlich mehr schematypische Intrusionen auftauchen als nach einer Woche. Der diesem Befund zugrunde liegende Mechanismus kann an dieser Stelle nicht hinreichend geklärt werden.

Betrachtet man die Dia- bzw. Unfall-Serien 'PKW-PKW' und 'PKW-Fahrrad' für beide Versuchsgruppen einzeln, so finden sich im Vergleich von T1 mit NT2 jeweils keine wesentlichen Unterschiede hinsichtlich konfabulierter Inhalte. Die Anzahl an Konfabulationen scheint demnach bei einer ersten Befragung und genanntem Vergleich nach einer bzw. acht Wochen zu keinem der beiden Zeitpunkte deutlich höher oder geringer zu sein. Lediglich über beide Dia-Serien hinweg betrachtet finden sich in den Schilderungen der Jugendlichen, die zwei Dia-Serien sahen merklich höhere entsprechende Fehlerraten, wenn diese erstmals nach acht Wochen befragt wurden.

Vergleich T2-NT2: Gruppe 'source monitoring':

Versuchspersonen der ESM – Gruppe scheinen hinsichtlich *beider Dia-Serien*, wenn sie nach acht Wochen bereits ein zweites mal befragt werden (T2), wesentlich besser und deutlich mehr Details korrekt erinnern zu können als wenn sie nach dem selben Zeitraum ein erstes Mal interviewt wurden (NT2). Dies gilt auch statistisch für die Gesamtzahl als auch für die im freien Bericht erinnerten korrekten Details. Tendenziell gilt dies nur für auf Nachfrage erinnerte Details.

Dieses klare Bild kann in diesem Zusammenhang offenbar nicht auf die Gesamtfehlerrate und die Zahl der allgemeinen Erinnerungsfehler übertragen werden. Hier macht es demzufolge keinen Unterschied, ob nach acht Wochen eine erste oder schon eine zweite Befragung stattfindet. Zwischen diesen Befragungszeitpunkten lassen sich keine signifikanten Unterschiede feststellen – zumindest nicht bei gleichzeitiger Betrachtung beider Dia-Serien. Ebenso wenig finden sich bedeutsame Unterschiede zwischen T2 und NT2 hinsichtlich von Quellenverwechslungen und auch Konfabulationen.

Bemerkenswerterweise kommen schematypische Intrusionen bei wiederholter Befragung nach acht Wochen deutlich häufiger vor als bei einmaliger Befragung ebenfalls nach acht Wochen.

Hinsichtlich getrennter Analyse der Dia-Serien ergab sich in der externalen 'source-monitoring' - Gruppe für die *Dia-Serie 'PKW-PKW'* außer bei den korrekt berichteten Details im freien Bericht für keine der anderen Variablen ein statistisch bedeutsamer Unterschied zwischen T2 und NT2. Innerhalb dieser Dia-Serie scheint es offensichtlich egal zu sein, ob

die Versuchspersonen nach acht Wochen schon einmal zuvor befragt wurden oder nicht. Bezüglich der korrekt erinnerten Details lässt sich zu T2 tendenziell eine höhere Erinnerungsrate konstatieren als zu NT2. Über alle Arten von Fehlern lassen sich kaum fundierte Aussagen treffen.

Für die *Dia-Serie 'PKW-Fahrrad'* zeigt sich in dieser Versuchsgruppe hinsichtlich der korrekt berichteten Details insgesamt und im freien Bericht ein anderes Bild. Hypothesenkonform erinnern die Jugendlichen bei einer zweiten Befragung (T2) nach acht Wochen wesentlich mehr Details als bei einer ersten Befragung nach acht Wochen (NT2). Tendenziell, jedoch statistisch nicht belegt, scheint dies auch für die auf Nachfrage korrekt erinnerten Details zu gelten.

Auch als nicht statistisch bedeutsam erweisen sich die Unterschiede zwischen diesen beiden Befragungszeitpunkten bezogen auf die zweite präsentierte *Dia-Serie 'PKW-Fahrrad'* für die Gesamtfehlerzahl, die Zahl der allgemeinen Erinnerungsfehler, der Arten von Quellenverwechslungen und Konfabulationen.

Lediglich hinsichtlich schematischer Intrusionen scheint hier von Bedeutung zu sein, ob die Versuchspersonen nach acht Wochen zum ersten Mal oder zum zweiten Mal befragt wurden. Bei einer zweiten Befragung finden sich hochsignifikant mehr Intrusionen als bei einer ersten.

Vergleich T2-NT2: Gruppe 'reality monitoring':

Im Gegensatz den Versuchspersonen der Bedingung 'external-external' erinnern die Jugendlichen dieser Versuchsbedingung *bezüglich beider Unfall-Serien* offenbar genau so gut oder schlecht korrekte Details zu T2 wie zu NT2, sei es insgesamt, im 'free recall' oder auf Nachfrage. Statistisch gesehen ergaben sich diesbezüglich zwischen den Befragungszeitpunkten keinerlei Unterschiede.

Ebenso umgekehrt zur ESM – Gruppe verhält es sich bezüglich der Unterschiede zwischen T2 und NT2 für die insgesamt gemachten Fehler sowie die allgemeinen Erinnerungsfehler. Bei erstmaliger Befragung nach acht Wochen treten hier für beide Fehlervariablen – betrachtet man beide Unfall-Serien simultan – sehr viel mehr Fehler auf als zu T2 bei bereits vorangegangener Befragung.

Wie schon in den voran stehenden Darstellungen auch ergaben sich auch hier beim Vergleich der begangenen Quellenverwechslungen und dem Auftreten von Konfabulationen keine relevanten Unterschiede zwischen den Befragungszeitpunkten.

Von Bedeutung sind jedoch auch hier wieder einmal schematische Intrusionen, welche in dieser Versuchsbedingung über beide Unfall-Serien hinweg bei einer ersten Befragung nach

acht Wochen (NT2) viermal häufiger genannt werden als bei einer wiederholten Befragung nach acht Wochen (T2).

Bei wiederum getrennter Betrachtung der Unfall-Serien ergab sich zunächst für die *Dia-Serie* 'PKW-PKW' lediglich für die Gesamtfehlerzahl sowie die schematypischen Intrusionen ein signifikanter Unterschied zwischen T2 und NT2. In beiden Fällen traten hier nach acht Wochen zu NT2 deutlich mehr Fehler auf als zu T2 mit einer vorausgegangenen Befragung. Hinsichtlich aller übrigen Fehlerarten sowie der korrekt berichteten Details scheint es in der Bedingung 'external-internal' im statistischen Ergebnis nicht so sehr von Bedeutung zu sein, ob nach acht Wochen eine erste oder eine zweite Befragung stattfindet. Tendenziell ergaben sich jedoch hypothesenkonforme Ergebnisse in Form von zu NT2 höheren Fehlerwerten und zu T2 höheren Werten bezüglich der korrekt berichteten Details.

Innerhalb der *selbst erfundenen Unfall-Serie* 'PKW-Fahrrad' fanden sich nur bei drei der zehn abhängigen Variablen deutliche Unterschiede zwischen T2 und NT2, nämlich hinsichtlich der Gesamtfehlerzahl sowie der Rate der allgemeinen Erinnerungsfehler (wie auch in der ESM – Gruppe) als auch hinsichtlich der schematypischen Intrusionen. Diejenigen Versuchspersonen, welche nach acht Wochen zum ersten Mal interviewt wurden, machten dann jeweils deutlich mehr Fehler in den genannten Formen als diejenigen, die sich zu diesem Zeitpunkt schon einer Befragung unterzogen hatten.

Für die übrigen Fehlerarten sowie die korrekt berichteten Details ließen sich jeweils nur unwesentliche Unterschiede zwischen T2 und NT2 nachweisen. Diesbezüglich scheint für die Gruppe 'external-internal' die Anzahl der Befragungen nach acht Wochen in diesem Rahmen von geringerer Bedeutung zu sein.

Die vorliegende Arbeit hatte des Weiteren das Ziel, zu untersuchen, wie sich Vernehmungen, die der aktuellen Befragung vorausgehen, auf diese auswirken, also die Frage, ob und in welchem Ausmaß bzw. in welchen Bereichen eine zweite Befragung zum selben Zeitpunkt wie eine erste Befragung vergleichbar oder von Vorteil bzw. Nachteil ist. Im Folgenden sollen also diejenigen Befunde des Vergleichs der Befragungszeitpunkte T2 und NT2 (jeweils nach acht Wochen) für beide Versuchsgruppen diskutiert werden:

Betrachtet man zunächst im Rahmen der Analyse der korrekt berichteten Details die Befunde des Vergleichs T2 gegenüber NT2 innerhalb der Gruppe der Jugendlichen, die zwischen einer externalen und einer internalen Quelle differenzieren sollten, so stellt man fest, dass sich hier weder für die gesehene Dia-Serie noch für die selbst erfundene Unfall-Serie bedeutsame Unterschiede zu den genannten Zeitpunkten feststellen haben lassen. Zwar berichten diese Jugendlichen bei einer zweiten Befragung nach acht Wochen über beide

Unfall-Serien hinweg wie auch jeweils einzeln insgesamt, wie auch im freien Bericht und auf Nachfrage erwartungsgemäß, aber eben nur tendenziell mehr korrekte Details als bei einer ersten ebenfalls nach acht Wochen; jedoch bleibt dieses Ergebnis ohne statistische Relevanz.

Anders dagegen verhält es sich für die Versuchspersonen der Bedingung 'external-external', also diejenigen, die Zeugen zweier Dia-Serien waren. Zwar ergaben sich auf Nachfrage hin für keine der Dia-Serien wesentlich mehr oder weniger korrekt berichtete Details zu einem der beiden Befragungszeitpunkte - es macht hier offensichtlich keinen Unterschied, ob einer Befragung eine andere vorausgeht oder nicht – allerdings ergaben sich bezüglich der insgesamt sowie im 'free recall' berichteten Details durchaus deutliche Unterschiede. So werden im freien Bericht bei der Abfrage beider Dia-Serien bei einer zweiten Befragung nach acht Wochen wesentlich mehr korrekte Einzelheiten berichtet gegenüber einer ersten Befragung nach acht Wochen.

So lassen die eben dargestellten Befunde den Schluss zu, dass eher nur diejenigen Jugendlichen, die zwischen zwei externalen Quellen differenzieren mussten, von einer zweiten Befragung nach acht Wochen bei vorausgegangener erster profitieren, was das Erinnern bzw. Berichten korrekter Details betrifft im Vergleich zu einer ersten nach acht Wochen. Für die Jugendlichen der anderen Versuchsbedingung scheint dies weniger relevant zu sein.

Hinsichtlich der Fehleranalyse finden sich auch beim Vergleich der Zeitpunkte T2 und NT2 einige Parallelen zum Vergleich T1 und NT2:

Betrachtet man beispielsweise wiederum die Gesamtfehlerzahl sowie die Zahl der allgemeinen Erinnerungsfehler, so kann festgestellt werden, dass analog zum Vergleich T1-NT2 sich innerhalb der Versuchsbedingung 'external-external' die beiden Befragungszeitpunkte hinsichtlich der beiden Fehlerraten jeweils nach acht Wochen nicht wesentlich unterscheiden. Allerdings ergeben sich ebenfalls in Analogie zum Vergleich T1-NT2 für die Versuchsbedingung 'external-internal' deutliche Unterschiede für beide Unfall-Serien. So werden von denjenigen Jugendlichen, die sowohl eine Dia-Serie sahen und sich auch eine zweite selbst ausdachten (intern generierte Quelle) deutlich mehr Fehler insgesamt sowie im Bereich allgemeiner Erinnerungsfehler begangen, wenn diese nach acht Wochen ein erstes Mal befragt werden als wenn nach acht Wochen bereits eine zweite Befragung stattfindet. Zwar erweist sich dieser Befund hinsichtlich der allgemeinen Erinnerungsfehler für die Dia-Serie 'PKW-PKW' statistisch nicht als signifikant, jedoch besteht bei genauerer Betrachtung der entsprechenden Mittelwerte auch hier eine deutliche Tendenz in die besagte Richtung.

Was das Vorkommen bzw. Unterschiede im Auftreten von Quellenverwechslungen - sei es im Vergleich von verschiedenen Befragungszeitpunkten oder auch generell zwischen den

Versuchsgruppen - angeht, ist die Befundlage auch hinsichtlich des letzten Vergleichs von T2 gegenüber NT2 bekannt und eindeutig. Wiederum unterscheiden sich auch eine erste Befragung nach acht Wochen (NT2) nicht von einer zweiten Befragung ebenfalls nach acht Wochen (T2). Sowohl Jugendliche, die zwischen zwei externalen Quellen differenzieren mussten, wie auch Jugendliche, die zwischen einer externalen und einer internalen Quelle unterscheiden mussten, begehen hinsichtlich keiner der beiden Dia- bzw. Unfall-Serien zu keinem der beiden Befragungszeitpunkte deutlich mehr oder weniger Quellenverwechslungen zwischen oder innerhalb der Serien. Über die zahlreichen, komplexen und differenzierten Ergebnisse der vorliegenden Arbeit scheint sich dieses Befundmuster als sehr stabil und konstant zu erweisen. Offensichtlich und interessanterweise begehen Jugendliche mit kognitiven Beeinträchtigungen insgesamt sehr wenig derartige Fehler unabhängig zwischen welchen Quellenarten sie unterscheiden müssen und unabhängig vom Befragungszeitpunkt. Interessant und aufschlussreich vor allem für die forensische Praxis wären weitere empirische Untersuchungen, die sich speziell und intensiv in verschiedensten Untersuchungsdesigns mit den konkreten Mechanismen der Quellenidentifikation bei jugendlichen Lernbehinderten über zeitlich größere und mehrere als zwei Befragungszeitpunkte befassen, da hier sicherlich gerade bei dieser spezifischen Personengruppe noch immenser Forschungsbedarf besteht.

Bezüglich konfabulierter Inhalte ist - zumindest in der vorliegenden Untersuchung - ein ebenfalls eindeutiges Befundmuster zu finden, was den Vergleich einer ersten Befragung nach acht Wochen mit einer zweiten zum selben Zeitpunkt angeht. Es zeigte sich, dass die Jugendlichen beider Versuchsgruppen weder bei Abfrage der beiden Dia- bzw. der Dia- und der Unfall-Serie zu einem Zeitpunkt mehr oder weniger konfabulierte Inhalte vorbringen als zu dem anderen Zeitpunkt. Offensichtlich macht es auch diesbezüglich keinen zu großen Unterschied, ob die Jugendlichen nach Wochen bereits eine Befragung absolviert hatten oder nicht. Lerneffekte aus vorangegangenen Befragungen sind demnach nicht so sehr zu erwarten gewesen.

Völlig anders und auch deshalb sehr interessant gestalten sich die Beobachtungen bezüglich des Vorkommens schematypischer Intrusionen zu den beiden Zeitpunkten. Hier ergaben sich in den beiden Versuchsgruppen völlig gegensätzliche Befunde – und zwar jeweils analog bzw. konstant über die beiden jeweiligen Dia- und Dia- bzw. Dia- und Unfall-Serien. Während diejenigen Jugendlichen, die zwischen einer externalen und einer internal selbst generierten Quellen unterscheiden mussten, dann wesentlich mehr Intrusionen produzierten, wenn sie nach acht Wochen ein erstes Mal zu den Serien befragt wurden gegenüber einer zweiten Befragung nach acht Wochen, verhält sich dies in der Gruppe 'external-external',

also bei Unterscheidung zweier gesehener Dia-Serien gänzlich umgekehrt. Hier berichten die Jugendlichen in beiden Dia-Serien bei einer zweiten Befragung deutlich mehr schematypische Intrusionen als bei einer ersten nach demselben Zeitraum, wobei das Signifikanzniveau bezüglich der Dia-Serie 'PKW-PKW' hier verfehlt wird. Tendenziell jedoch stimmen die Ergebnisse überein.

Möglicherweise werden bestimmte Unfall-Schemata oder Skripte bei Beobachtung zweier Dia-Serien erst durch eine zweite Befragung aktiviert bzw. kann dann erst darauf zugegriffen werden. Somit scheint eine vorausgehende Befragung für die ESM – Gruppe eine Art sich diesbezüglich negativ auswirkenden Lerneffekt zu haben.

5.1.4 Diskussion zu den Interaktionseffekten

Hinsichtlich der Fragestellung, ob sich für die in den Aussagen berichteten korrekten Details sowie für die verschiedenen Fehlerarten Veränderungen in der Ausprägung in Abhängigkeit von der Gruppenzugehörigkeit vom ersten zum zweiten Befragungszeitpunkt ergeben, konnten lediglich für die abhängigen Variablen 'Fehler gesamt' wie auch 'allgemeine Erinnerungsfehler' statistisch signifikante Wechselwirkungen festgestellt werden.

Aufgrund der Befunde kann zumindest im Bereich der genannten Fehlervariablen, die anfangs aufgestellte Arbeitshypothese zugunsten einer Alternativhypothese verworfen werden. Offenkundig hat der Faktor Zeit doch einen negativeren Einfluss auf die Gedächtnisleistung - was die Gesamtfehlerzahl sowie die Zahl der allgemeinen Erinnerungsfehler angeht - wenn zwischen einer externalen und einer internalen Quelle unterschieden werden muss, als wenn zwischen zwei externalen Quellen differenziert werden muss.

Neben den dargestellten Ergebnissen zu den verschiedenen Haupteffekten 'Gruppe' und 'Befragungszeitpunkt' konnten für keine der übrigen Variablen statistisch bedeutsame Interaktionseffekte festgestellt werden.

Zwar ergaben sich bezüglich der Fehlerarten 'Intrusionen' und 'Konfabulationen' jeweils deutliche Unterschiede zwischen den Versuchsgruppen, dergestalt, dass diejenige Versuchsgruppe, die zwischen zwei externalen Quellen differenzieren musste, wesentlich mehr Fehler der Art 'schematypische Intrusionen' beging als diejenige, die zwischen einer externalen und einer internalen Quelle unterscheiden musste. Umgekehrt traten in der Gruppe, die zwischen einer externalen und einer internalen Quelle differenzieren sollte, deutlich mehr Konfabulationen auf als in der Bedingung 'external-external'. Zudem erweist sich hinsichtlich der Konfabulationen der Einfluss eines zweiten Befragungszeitpunktes insofern als bedeutsam, als dass im Laufe der Zeit derartige Fehler öfter begangen werden. Wechselseitige Abhängigkei-

ten konnten wie bereits erwähnt jedoch weder für Intrusionen noch für Konfabulationen beobachtet werden.

Neben der Tatsache, dass hinsichtlich begangener Quellenverwechslungen festgestellt werden kann, dass lediglich zwischen den Serien Jugendliche, die zwischen einer externalen und einer internalen Quelle differenzieren mussten, eine höhere Fehlerrate aufweisen, als diejenigen, die zwischen zwei externalen Quellen unterscheiden sollten (obwohl der gefundene Effekt bezüglich seiner Effektstärke als relativ gering anzusehen ist), konnten ansonsten im Rahmen Quellenverwechslungsfehler keinerlei sonstige Gruppenunterschiede, Effekte von Befragungszeitpunkten bzw. deren zeitlichem Abstand voneinander oder wechselseitige Abhängigkeiten gefunden werden.

Bemerkenswerterweise kommt dem gesamten Bereich der korrekt erinnerten oder zumindest korrekt berichteten Details eine geringere Bedeutung zu als angenommen, obwohl ein statistisch relevanter Gruppenunterschied im Bereich der auf Nachfrage korrekt berichteten Details nur sehr knapp verfehlt wird. Hierbei ist der Befund interessant, dass auf Nachfrage diejenigen Jugendlichen, die zwischen einer externalen und einer internalen Quelle unterscheiden mussten, mehr derartige Einzelheiten korrekt erinnern als die Jugendlichen der Bedingung external-external, was wiederum den tendenziellen Ergebnissen bezüglich der insgesamt sowie im free recall korrekt berichteten Details gegenüber steht. Hier deutet sich also an, dass gerade in solchen Quellenidentifikationskonstellationen wohl eine sorgsame und angemessene Interview- und Fragetechnik entscheidenden Anteil an einer Optimierung der Gedächtnisleistung haben kann. Im Übrigen konnte keine Interaktion von Gruppenzugehörigkeit und Befragungszeitpunkt beobachtet werden.

So scheinen zusammenfassend lediglich im Bereich der Gesamtfehlerzahl sowie der allgemeinen Erinnerungsfehler wechselseitige Abhängigkeiten von Gruppenzugehörigkeit und Befragungszeitpunkt zu bestehen. Alle übrigen Fehlerarten sowie der gesamte Aspekt korrekt berichteter Details kann bzw. konnte in dieser Untersuchung hinsichtlich der Frage wechselseitiger Interaktion als nicht besonders relevant betrachtet werden.

5.2 Diskussion zu den Befunden der Glaubhaftigkeit von Zeugenaussagen (Validierungsuntersuchung)

Da wie bereits erwähnt, bis dato keine entsprechenden Validierungsuntersuchungen der aussageimmanenten Qualitätsmerkmale (Realkennzeichen) bei Personen mit Intelligenzminderung vorliegt, kann in der folgenden Diskussion keine direkte Bezugnahme darauf erfolgen. Allerdings soll neben spezifischen möglichen Erklärungsmodellen und Diskussionspunkten bei bestimmten Aspekten dennoch versucht werden, eine Einordnung bzw. einen Vergleich mit bisher vorliegenden Befunden zu ziehen bzw. auf nahe liegende Hintergründe verweisen.

5.2.1 Diskussion zu den Befunden der einzelnen Aussagemerkmale

In zahlreichen Validierungsuntersuchungen sowohl in erster Linie bei Kindern als auch bei Erwachsenen hat sich herausgestellt, dass das Qualitätsmerkmal *‘Logische Konsistenz’* einen sehr guten Prädiktor zur Differenzierung wahrer von erfundenen Aussagen darstellt. Dies bestätigt sich auch in der vorliegenden Arbeit. Die statistischen Analysen ergaben deutliche Unterschiede in den Merkmalsausprägungen. In den Geschichten ohne Erlebnishintergrund lag zu beiden Befragungszeitpunkten die Ausprägung dieses Aussagemerkmals deutlich unter der von tatsächlich erlebten Geschichten. Lernbehinderten Jugendlichen fällt es sichtlich schwer logisch nachvollziehbare und schlüssige Aussagen zu konstruieren und glaubhaft zu reproduzieren, wenn dafür keine Erlebnisgrundlage vorhanden ist. Die Ursachen können wie ausführlich dargestellt in den verschiedensten kognitiven Fähigkeiten und Bereichen liegen. Allerdings konnte weder für die wahren noch für die erfundenen Geschichten eine merkliche Veränderung vom ersten zum zweiten Befragungszeitpunkt beobachtet werden. Der Zeitpunkt der Befragung bzw. der zeitliche Abstand zwischen zwei Befragungen scheint in dieser Untersuchung keine so große Rolle zu spielen.

Ein ebenso hypothesenkonformes Ergebnis findet sich analog zum Qualitätsmerkmal *‘Logische Konsistenz’* auch für das im aussagepsychologischen Begutachtungskontext ebenfalls sehr bedeutsame Kriterium der *‘widerspruchsfreien, ungeordneten Darstellung’*. Auch diesbezüglich haben die kognitiv beeinträchtigten Jugendlichen Schwierigkeiten, nicht selbst erlebte Geschichten in der „geforderten“, die Glaubhaftigkeit der Aussage unterstützenden Art und Weise entsprechend zu schildern, was kognitiv wohl eine über den Kompetenzen der Jugendlichen stehende Anforderung ist. So trennt dieses Aussagemerkmal unabhängig vom Befragungszeitpunkt, d.h. sowohl zu T1 als auch zu T2, sehr deutlich

zwischen erlebnisfundierte und nicht erlebnisfundierte Aussagen. Eine signifikante Veränderung über die Zeit hinweg konnte jedoch auch hier für beide Aussagearten nicht festgestellt werden. Zwar findet sich für die Geschichten mit Erlebnishintergrund eine leichte tendenzielle, hypothesenkonforme Entwicklung im Sinne einer Abnahme der Ausprägung des Merkmals vom ersten zum zweiten Befragungszeitpunkt, jedoch ist für die erfundenen Geschichten eine umgekehrte Tendenz festzustellen gewesen.

Das dritte Merkmal - der *'Quantitative Detailreichtum'* - der besonders wichtigen und bedeutsamen ersten Kategorie der *'Allgemeinen Merkmale'* bestätigt ebenfalls bekannte Befunde verschiedener Simulations- und Feldstudien und weist in dieser Untersuchung hypothesenkonforme Befunde auf. Offensichtlich sind Jugendliche mit intellektuellen Defiziten weniger gut in der Lage, Geschichten ohne Erlebnishintergrund „künstlich“ mit Details auszuschnürcen bzw. diese zu erinnern. Auch dieses inhaltliche Qualitätsmerkmal erweist sich in dieser Validierungsuntersuchung als trennscharfes Kriterium, um zwischen wahren und erfundenen Aussagen zu differenzieren. Entsprechende Merkmalsausprägungen fallen in erlebnisfundierte Aussagen in Form eines höchstsignifikanten Haupteffekts *'Wahrheitsstatus'* deutlich höher aus als in nicht erlebnisfundierte. Post-hoc-analytisch zeigen sich für beide Befragungszeitpunkte entsprechende Ergebnisse. Der zeitliche Aspekt ist allerdings auch erneut weniger relevant; obwohl sich für beide Geschichten (wahr und erfunden) im Mittel und hypothesengemäß zum zweiten Befragungszeitpunkt zwar geringere Merkmalsausprägungen finden als zum ersten, sind diese jedoch nur tendenziell zu verzeichnen. Die aufgrund der in manchen Untersuchungen festgestellten Befunde von zu einem weiteren Befragungszeitpunkt höheren Ausprägungsgrad des Detailreichtums im Sinne der erläuterten Phänomene *'Reminiszenz'* oder *'Hypermnesie'* konnten demnach zumindest in dieser Teiluntersuchung nicht bestätigt oder als Erklärung herangezogen werden. Zwar waren die Jugendlichen beim zweiten Interviewtermin und auch aufgrund der anderen, vorangegangenen Teiluntersuchung (source-/ reality monitoring) bereits an die Befragungssituation gewöhnt und wussten in etwa was und wer auf sie zukommt („Test-Sophistication“), jedoch fand dies zumindest im Rahmen der *'Allgemeinen Merkmale'* noch keinen Niederschlag.

Insgesamt können die ersten drei Qualitätsmerkmale offensichtlich auch bezüglich der Population lernbehinderter Jugendlicher als sehr trennscharfe Kriterien betrachtet werden, die ein hohes Differenzierungspotential aufweisen. Hinsichtlich der zeitlichen Dimension kann festgehalten, dass die verwendeten Kriterien der ersten Kategorie zu beiden Befragungszeitpunkten, also nach einer Woche in etwa vergleichbare oder ähnliche diskriminative Eigenschaften haben wie nach ca. sechs Wochen.

Das bisher beobachtete Befundmuster setzt sich offensichtlich auch in der zweiten Merkmalskategorie der 'Speziellen Inhalte' fort.

Das aufgrund der Ergebnisse zahlreicher Simulations- und Validierungsuntersuchungen besonders für die aussagepsychologische Begutachtung (gerade in Fällen des sexuellen Missbrauchs) ebenfalls recht bedeutsame Qualitätsmerkmal der '*Kontextuellen Einbettung*' scheint auch in dieser Studie ein verlässliches und gutes Diskriminationsinstrument zu sein, da sich hier zumindest wiederum zu beiden Befragungszeitpunkten jeweils hochsignifikante Unterschiede zwischen den Versuchsgruppen ergaben. Dieser Befund deckt sich mit fast allen Ergebnissen vorangegangener Untersuchungen bei Kindern (vgl. LAMERS-WINKELMAN & BUFFING, 1996; ANSON et al., 1993; LAMB, STERNBERG et al., 1997; TYE et al., 1999; STELLER et al., 1992 und im besonderen HOROWITZ, LAMB et al., 1997), nach deren Ergebnissen dieses Realkennzeichen sehr häufig in den Aussagen beider Befragungszeitpunkte zu finden war. Wiederum konnte zwar sowohl für erlebnisbasierte als auch für erfundene Aussagen hypothesenkonform eine leichte tendenzielle Verringerung der Merkmalsausprägung vom ersten zum zweiten Befragungszeitpunkt beobachtet werden, die jedoch statistisch unbedeutend blieb.

Schildert ein Zeuge *nonverbale Aktionen*, gilt dies in der aussageanalytischen Zeugenpsychologie ebenfalls als spezieller Aussageinhalt mit einer guten Belegkraft hinsichtlich erlebnisfundierter Geschichten. In der vorliegenden Untersuchung stellt sich dieses Aussagemerkmal als besonders relevant dar, da es zum einen höchstsignifikant zwischen wahren und erfundenen Aussagen trennt, zum anderen sich auch ein zeitlich sehr bedeutsamer Effekt gemäß den Erwartungen (Hypothesen) finden ließ.

Die genauere post-hoc-Analyse ergab zu beiden Befragungszeitpunkten eine hohe Diskriminationsfähigkeit zwischen den beiden Aussagearten. Auch in Aussagen lernbehinderter Jugendlicher finden sich in erfundenen Geschichten deutlich weniger derartige Schilderungen.

Zudem scheinen die Jugendlichen Probleme damit zu haben, dieses Qualitätsmerkmal sowohl bei einem Erlebnishintergrund als auch ohne einen solchen über die Zeit hinweg konstant in weitere, nachfolgende Schilderungen einzubauen bzw. zu erinnern. Dieser Befund einer deutlichen Abnahme der Merkmalsausprägung von T1 nach T2 in beiden Aussagen findet sich in dieser Untersuchung ebenfalls lediglich bei dem Merkmal der 'Schilderung eigener psychischer oder physischer Vorgänge'.

Nach GREUEL (2001) haben neuere Feldstudien gezeigt, dass „die Abfolge von Interaktions- bzw. Ereignissequenzen sehr schnell vergessen wird, hinsichtlich der Angaben zur Chronologie komplexer Abläufe bereits nach wenigen Tagen mit unzuverlässigen Angaben gerechnet werden muss“ (GREUEL, 2001, S. 124). Die Untersuchung von GORDON et al.

(1994) zeigte dennoch, dass auch Kindern mit einer Intelligenzminderung die Erinnerung an selbst ausgeführte Handlungen leichter fiel als an solche, die nur vorgestellt waren. Die Erinnerungsfähigkeit nahm jedoch nach sechs Wochen wie auch in dieser Studie insgesamt ab (vgl. GORDON et al., 1994). Trotzdem die erlebnisbasierten Ereignisse im Durchschnitt wesentlich länger als sechs Wochen zurücklagen, fand sich in den erlebnisfundierten Aussagen dennoch eine deutlich höhere Merkmalsausprägung, die bzw. deren Erinnerung an selbst ausgeführte Handlungen bzw. beobachtete Handlungen wie auch hinsichtlich erfundener Geschichten nach relativ kurzer Zeit bereits verblasst ist.

Bei der *Wiedergabe von Gesprächen* konnte insgesamt betrachtet zwar kein statistisch bedeutsamer Haupteffekt des Wahrheitsgehalts festgestellt werden, dafür aber hinsichtlich des Befragungszeitpunktes. Zwar fanden sich bei genauerer Betrachtung tendenziell zu beiden Befragungszeitpunkten in erlebnisbasierten Aussagen höhere Mittelwertsausprägungen als in solchen ohne Erlebnishintergrund, jedoch erwies sich dieser Unterschied lediglich zum zweiten Befragungszeitpunkt als statistisch bedeutsam. Zumindest die tendenziellen Befunde stehen in Einklang mit fast allen referierten Simulationsstudien bei Kindern, selbst mit STELLER et al. (1992), wo dieses Merkmal das Signifikanzniveau wie auch zum Teil in der vorliegenden Untersuchung nur knapp verfehlt wird. Insgesamt lassen sich scheinbar auch bei Jugendlichen mit kognitiven Beeinträchtigungen Tendenzen in die hypothesenkonforme Richtung feststellen. Nach GREUEL (2001) kann im Übrigen der genaue Wortlaut von Gesprächen nur außerordentlich schlecht behalten werden. Die Erinnerungsleisten seien nur dann besser, wenn die Äußerungen Aufforderungscharakter hatten oder für den Rezipienten überraschend waren. Hingegen wird „der Kerninhalt von Gesprächen (gist memory), insbesondere aber der Bedeutungsgehalt (thematic recall) auch über längere Erinnerungsintervalle zuverlässig erinnert“ (GREUEL, 2001, S. 126). Dies gilt insbesondere für Schilderungen ohne Erlebnishintergrund. Zu prüfen bleibt inwieweit Jugendliche mit einer Intelligenzminderung in der Lage sind der komplexen Aufgabe gerecht zu werden, differenzierte Abfolgen von Rede und Gegenrede zu schildern bzw. zu erinnern.

Weiterhin lassen die Befunde der post-hoc-Analyse den Schluss zu, dass es den Jugendlichen bei ihrer Schilderung der erfundenen Geschichten noch schwerer fällt, entsprechende erdachte Gesprächsinhalte über die Zeit hinweg (von T1 nach T2) aufrechtzuerhalten als bei wahrheitsgemäßen Berichten. Es fand sich lediglich bei den erfundenen Aussagen eine deutliche Abnahme des Aussagemerkmals vom ersten zum zweiten Befragungszeitpunkt.

Demnach gilt dieses Realkennzeichen relativ unbestritten als sehr trennscharf, da aus den dargestellten gedächtnispsychologischen Befunden und Erkenntnissen der Informationsverarbeitung die an dieses Kriterium gestellten Anforderungen im Sinne von Perspektivenwechsel, Rollenübernahme und logischer Kombinationsgabe von einem intentional falsch aussa-

genden Zeugen, insbesondere wenn dieser intellektuell eingeschränkt ist, kaum in schlüssiger Art und Weise erfüllt werden können und dessen Planungskapazität übersteigen. Zudem muss bezüglich dieser Untersuchung allerdings festgestellt werden, dass sich die Ausprägungen dieses Qualitätsmerkmals insgesamt in einem relativ geringen Auftretensspektrum bewegen, also sowohl in wahren wie in erfundenen Geschichten nicht besonders häufig vorkommen. Insgesamt kann auch in dieser Untersuchung bei der speziellen Personengruppe der lernbehinderten Jugendlichen dem Realkennzeichen ein gewisses Potential zugesprochen werden, da sowohl hinsichtlich des Wahrheitsstatus als auch des Befragungszeitpunktes zumindest tendenziell hypothesenkonforme und den Vermutungen entsprechende Ergebnisse zu beobachten sind.

Zwar konnten für das Merkmal der *‘Schilderung von Komplikationen im Handlungsverlauf’* sowohl ein signifikanter Haupteffekt des Wahrheitsstatus wie auch des Befragungszeitpunktes gefunden werden, allerdings trennen derartige Schilderungen nur zum zweiten Befragungszeitpunkt signifikant zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen. Hier scheint also eine zeitliche Verzögerung bzw. ein zweiter Befragungstermin für lernbehinderte Jugendliche von Bedeutung zu sein, obwohl wiederum nur tendenziell zu beiden Befragungszeitpunkten erwartungsgemäß in erlebnisbasierten Aussagen mehr Schilderungen solcher Komplikationen bzw. Aussagemerkmale zu finden sind.

Ein ebenfalls interessanter, differenzierter Befund ergab sich hinsichtlich der Veränderung über die Zeit hinweg. Hierbei ließ sich zwar erneut tendenziell für beide Geschichten eine Abnahme der Merkmalsausprägung verzeichnen, jedoch war diese zeitliche Veränderung nur für die erfundenen Geschichten statistisch relevant. Vergessensprozesse bzw. eine mangelnde Fähigkeit, die ursprünglich so erdachten Aussageinhalte zu einem späteren Zeitpunkt wieder in ähnlichem Umfang zu reproduzieren, scheinen somit für Jugendliche mit einer Intelligenzminderung hinsichtlich frei erfundener Geschichten erheblich problematischer zu sein als hinsichtlich tatsächlich erlebter autobiographischer Ereignisse.

Insgesamt ergab sich dennoch wiederum eine grundsätzliche Übereinstimmung mit den Befunden zitierter Studien. Allerdings ist zu bemerken, dass gerade in Simulationsstudien das Vorkommen solcher Aussagemerkmale (wie auch aller Merkmale der Kategorien *‘Spezielle Inhalte’* und *‘Inhaltliche Besonderheiten’*) verstärkt von der Auswahl, Qualität, Struktur und Intensität der Themen der autobiographischen Ereignisse wie auch der erfundenen Themen abhängt.

Auch hinsichtlich der Schilderung *‘ausgefallener Details’* ergaben sich recht differenzierte Befunde. Vorweg muss allerdings angemerkt werden, dass solche Inhalte nur sehr selten zu verzeichnen waren und geratet werden konnten. Das Vorkommen bestimmter Merkmale (vor

allem aus den Kategorien 'Spezielle Inhalte' und besonders 'Inhaltliche Besonderheiten') ist natürlich immer davon abhängig, ob diese im Originalerlebnis auch tatsächlich stattgefunden haben. Dies ist beim Rating von Aussagen stets zu beachten, weswegen in dieser Untersuchung generell eine sehr geringe Ausprägung dieses Merkmals in den Aussagen beider Versuchgruppen zu beiden Befragungszeitpunkten zu finden war. So konnte das Merkmal analog zu obigen Ausführungen der Schilderung von Komplikationen im Handlungsverlauf zum ersten Befragungszeitpunkt nicht zwischen wahren und falschen Aussagen trennen, jedoch sehr deutlich zum zweiten. Auch hier sind die Jugendlichen offenkundig nicht oder nur schwer in der Lage, an einem weiteren Befragungstermin entsprechende Aussageelemente in ihre „Falschaussage“ einzubauen.

Über die beiden Befragungszeitpunkte hinweg ließ sich wiederum analog zum Kriterium 'Schilderung von Komplikationen im Handlungsverlauf' lediglich für diejenigen Aussagen ohne Erlebnishintergrund eine deutliche Veränderung vom ersten zum zweiten Befragungszeitpunkt konstatieren.

In diesem Zusammenhang sei auf die Untersuchung von KRAHÉ & KUNDROTAS (1992) verwiesen, die in einem aussageanalytischen Feldexperiment feststellten, dass das Kriterium 'ausgefallene Details' signifikant häufiger in Falschaussagen von Erwachsenen gefunden wurde. Gerade bei den gewählten Themen dieser Untersuchung, die in der Regel weniger komplex sind als ein Missbrauchserlebnis, besteht die Möglichkeit, dass die Geschichten der Jugendlichen einfach kaum außergewöhnliche Einzelheiten aufweisen. Zum anderen stellt sich die Frage, ob ausgefallene Einzelheiten, die nicht schemakonsistent sind, von den Jugendlichen vorgetragen wurden oder wegen ihrer Außergewöhnlichkeit nicht für einen späteren Abruf kodiert wurden.

Weiterhin bedeutsam für die aussagepsychologische Begutachtung von jugendlichen Lernbehinderten erscheint auf Grundlage der Ergebnisse der vorliegenden Arbeit auch das Glaubhaftigkeitsmerkmal der '*Schilderung nebensächlicher Details*'. Zum einen belegen die Ergebnisse für beide Befragungszeitpunkte deutlich und statistisch relevante höhere Merkmalsausprägungen in wahren als in erfundenen Geschichten. Dies spricht für die Annahme, dass Personen mit kognitiven Einschränkungen nicht über die erforderliche Kapazität verfügen, bei der Schilderung von Aussagen ohne tatsächlichen Erlebnishintergrund diese noch mit zusätzlichen, nicht so sehr bedeutungsvollen Einzelheiten auszus schmücken bzw. diese in einer Schilderung entsprechend darzustellen. Zwar machten einige der Jugendlichen im Rahmen der Befragungen kommunikativ einen relativ guten bzw. „fitten“ Eindruck, da sie ja auch zudem entsprechend instruiert wurden, dem Interviewer (= Verfasser der vorliegenden Arbeit) die erfundenen Geschichten (im Rahmen ihrer Kompetenzen) möglichst glaubhaft zu schildern, jedoch wurden während der Auswertung der Interviews (Rating) schnell entspre-

chende Defizite und Probleme offenkundig. Dabei besteht auch die Möglichkeit, dass die Jugendlichen mit einer Intelligenzminderung weniger zwischen relevanten und irrelevanten Details zu unterscheiden wussten.

Zum anderen konnte nur für die erlebnisfundierten Geschichten eine signifikante Veränderung im Sinne einer Abnahme des Vorkommens des Merkmals von T1 nach T2 festgestellt werden.

Die genannten Befunde sprechen jedoch im Vergleich mit anderen Untersuchungen trotzdem für eine kritische Reliabilität, da das Kriterium je nach Versuchsmaterial, Untersuchungsdesign oder dem intellektuellen Niveau bzw. der kognitiven Kapazität der Versuchspersonen knapp zwischen den Versuchsgruppen trennt oder das jeweilige Signifikanzniveau um wenig verfehlt. In einigen Befunden aus anderen Untersuchungen etwa (z.B. LAMERS-WINKELMAN, 1996; LAMB et al., 1997; HOROWITZ et al., 1997) wies dieses Merkmal keine allzu große Bedeutung auf, zwischen wahren und falschen Aussagen differenzieren zu können. Im Gegensatz dazu trug es in der Simulationsstudie von STELLER et al. (1992) durchaus zur Trennung zwischen den Versuchsgruppen bei. In der Untersuchung von ANSON et al. (1993) fand sich im Ergebnis eine mittlere, also grenzwertige Reliabilität. Das Ergebnis dieser Studie ist zumindest hinsichtlich des Befragungszeitpunktes insofern kritisch zu betrachten, da das Merkmal wie dargestellt nur für die erlebnisfundierten Aussagen eine Veränderung zwischen den Befragungszeitpunkten aufweist. Die Schilderung derartiger Einzelheiten in erfundenen Aussagen scheint von zeitlichen Veränderungen nicht so sehr betroffen.

Die Aussagemerkmale *‘phänomengemäße Beschreibungen unverstandener Handlungselemente’* und *‘Inhaltliche Verschachtelungen’* konnten sowohl in wahren wie auch in erfundenen Aussagen so gut wie nicht festgestellt werden und wurden somit aufgrund dieses „bottom-effects“ auch nicht in die statistische Analyse miteinbezogen. Nach KÖHNKEN (2001) kommt phänomengemäßen Beschreibungen besonders in Aussagen von Kindern eine hohe Belegkraft zu und sind in solchen vermehrt zu finden, da Kinder noch einen geringeren Fundus an Schemata haben. Insbesondere dann, wenn es sich um sexuelle Praktiken und physische Vorgänge handelt, schildern Kinder dann des Öfteren die Vorgänge nur aufgrund ihrer äußeren Anmutqualität (vgl. KÖHNKEN, 2001). Die Jugendlichen in dieser Studie berichteten jedoch eher über alltägliche Erlebnisse, oder zumindest über solche bei denen die Versuchspersonen über ein Schema verfügen sollten, da solche Erlebnisse durch die Medien hinreichend bekannt sind (z.B. Krankenhausaufenthalt mit Operation). Aus diesem Grund wurden vermutlich kaum phänomengemäße Beschreibungen in den Aussagen gefunden und wenn, dann nur in den erlebnisbasierten Aussagen, was für die Schwierigkeit spricht, solche Schilderungen zu konstruieren.

Auch die *'indirekt Handlungsbezogene Schilderungen'* wurden aufgrund eines Bodeneffektes nicht ausgewertet, da sie ebenfalls faktisch quasi nicht gefunden wurden. Nach ARNTZEN (1993) handelt es sich bei diesem Realkennzeichen auch um eine seltenere Komplikationsform. Auch wenn es ein trennscharfes Realitätskriterium zu sein scheint, wie verschiedene Studien belegen, kann es in dieser Validierungsuntersuchung aufgrund der thematischen Vorgaben, aber auch des Alters und (zwar eingeschränkten) wissensbasierten Hintergrundes der Versuchspersonen keinen Beitrag zur Diskrimination zwischen erlebnisbasierten und nicht-erlebnisbasierten Aussagen liefern (vgl. GREUEL, 2001).

Besonders hoch einzuschätzen ist in dieser Untersuchung der diskriminatorische Wert des Qualitätsmerkmals der *'Schilderung eigener psychischer oder physischer Vorgänge'*. Sowohl bezüglich beider Befragungszeitpunkte fanden sich in erlebnisfundierten Aussagen deutlich höhere Merkmalsausprägungen als in nicht erlebnisfundierten, als auch konnte für wahre wie erfundene Geschichten eine deutliche Verringerung der Mittelwerte des Realkennzeichens vom ersten zum zweiten Befragungszeitpunkt festgestellt werden. Es fanden sich also gemäß den Erwartungen große Unterschiede generell zwischen wahren und erfundenen Geschichten wie auch deutliche Veränderungen über die Zeit hinweg, was auch die Ergebnisse der Diskriminanzanalyse bestätigen, wonach dieses Qualitätsmerkmal zu beiden Zeitpunkten vordere Rangplätze hinsichtlich der diskriminatorisch bedeutsamen Merkmale einnimmt. Die Schilderung eigener psychischer Vorgänge schien den Jugendlichen, im Vergleich zu anderen inhaltlichen Qualitäten, relativ leicht zu fallen. Dieser Befund muss in differenzierter Weise in Bezug zu vorangegangenen Untersuchungsergebnissen gesetzt werden. Die Schilderung eigenpsychischer Vorgänge gilt nach den Ergebnissen von STELLER et al. (1992), LAMB et al. (1997) sowie ORBACH & LAMB (1999) als gutes und geeignetes Kriterium um zwischen erlebnisbegründeten und erfundenen Aussagen von Kindern zu unterscheiden, während es in anderen Studien wiederum eher geringere Trennkraft bewies. Im Zusammenhang mit der vorliegenden Untersuchung gilt es jedoch zu beachten, dass die Beschreibung eigener psychischer Vorgänge, wie jede Äußerung des Jugendlichen, aus der Erinnerung erfolgen muss. Derzeit wird davon ausgegangen, dass Erinnerungen an psychische Reaktionen, wie Gedanken und Emotionen oder Affekte, generell reflexiven Charakter haben, d.h. immer retrospektiv vom aktuellen Ist-Zustand rekonstruiert werden müssen. Daher unterliegen solche Erinnerungen auch Überformungen oder Verzerrungen (vgl. GREUEL, 2001).

Dagegen ergibt sich bezüglich der *'Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten'* ein gänzlich anderes Bild. Zum einen muss zunächst festgestellt werden, dass sich die Ausprägungen des Merkmals in beiden Geschichten und zu beiden

Befragungszeitpunkten innerhalb eines sehr geringen Spektrums bewegen, also nur in sehr geringer Qualität vorkommen. Demzufolge kommt dem Realkennzeichen zu keinem Befragungszeitpunkt eine große Trennschärfe zu, um zwischen wahren und erfundenen Aussagen differenzieren zu können. Zum anderen kann zwar tendenziell für erlebnis- wie für nicht erlebnisbasierte Aussagen eine leichte Abnahme der Merkmalsausprägung beobachtet werden, jedoch ohne statistische Relevanz. Die Schilderung psychischer Vorgänge beim Beschuldigten scheint an sich somit ein etwas indifferenter Prädiktor zu sein. Während er sich etwa bei ANSON et al. (1993) durch eine adäquate Reliabilität auszeichnete, kam man nach den Ergebnissen von LAMERS-WINKELMAN & BUFFING (1996), HOROWITZ et al. (1997) und LAMB et al. (1997) zu dem Schluss, dass dieses Kriterium doch nicht unbedingt als besonders trennscharf einzustufen sei.

Beide gerade dargestellten Qualitätsmerkmale scheinen zum einen vom tatsächlichen Geschehen bzw. dessen spezieller Thematik abhängig zu sein, zum anderen von der persönlichen Relevanz bzw. der Empathiefähigkeit der/des Zeugen. Nach den im theoretischen Teil dargestellten Hintergründen, scheint es für manche Jugendliche mit intellektuellen Einschränkungen kritisch und problematisch zu sein, sich ohne weiteres in soziale Situationen einzufügen und dort zurechtzukommen. Demzufolge treten entsprechende Schwierigkeiten auf im Sinne von notwendigen Rollenübernahmen oder Perspektivenwechseln. Die folgende mögliche Erklärung für die gefundenen Effekte beider Realkennzeichen steht eventuell wiederum in Zusammenhang mit der gesamten Menge der berichteten Details. Der Einfluss von Empathie auf die Menge der wiedergegebenen Information könnte beispielsweise dadurch zustande kommen, dass hochempathische Personen einer sozialen Situation mehr Aufmerksamkeit widmen als weniger empathische und es deshalb zu einer besseren Enkodierung der Information kommt oder diese Personen aufgrund ihres stärkeren Interesses für soziale Vorgänge in einer Befragungssituation ein *erhöhtes Mitteilungsbedürfnis* haben. Weiterhin könnten ein *erhöhtes emotionales 'arousal'* oder eine intensivere gedankliche Auseinandersetzung und Weiterbeschäftigung mit der sozialen Situation damit verbunden sein. Dass dies wiederum die Vergessensprozesse abschwächt, wäre damit erklärbar, warum bei empathischeren Personen die Gesamterinnerungsleistung steigt, wenn sie „empathieauslösende“ soziale Situationen erinnern.

Sind diese Prozesse wie bei vielen intellektuell eingeschränkten Personen gestört oder nur unzureichend entwickelt, ergeben sich möglicherweise entsprechende Defizite, psychische Vorgänge anderer zu erkennen oder zu verstehen. Umgekehrt bzw. demgegenüber scheinen behinderte Menschen aufgrund ihrer sozialen Situation und eventuellen Problemen im Umgang mit Mitmenschen auch eher auf sich und internale, emotionale Vorgänge fokussiert als auf externe und somit womöglich besser in der Lage, entsprechende Emotionen wahrzunehmen und zu schildern, wenn diese auf tatsächlichen Gegebenheiten basieren. Eine

fiktionale Konstruktion solcher Vorgänge und deren Reproduktion scheint sowohl bei auf das Selbst wie auch auf andere Personen bezogenen Prozessen zu komplex.

Im Rahmen der vierten Kategorie der Aussagemerkmale, den 'Motivationsbezogenen Inhalten' ergeben sich wie eigentlich auch in anderen referierten Untersuchungen bei Kindern eher indifferente Befunde. Um einen anderen Aspekt kurz vorwegzunehmen, kann festgehalten werden, dass bezüglich keines der Merkmale aus dieser Kategorie - ebenso wie für die wichtigste Kategorie der 'Allgemeinen Merkmale' - die Veränderung von einem Befragungszeitpunkt zu einem anderen von großer Bedeutung in Form eines statistischen Haupteffekts 'Befragungszeitpunkt' war.

In der vorliegenden Untersuchung ergab die post-hoc-Analyse der Ergebnisse zur '*spontanen Verbesserung der eigenen Aussage*', dass das Merkmal gut zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen differenzieren kann. Dies gilt insbesondere für den ersten Befragungszeitpunkt. Trotzdem für den zweiten Befragungszeitpunkt nur denkbarst knapp das Signifikanzniveau verfehlt wurde, kann deswegen auch für einen weiteren Befragungszeitpunkt von einer guten hypothesenkonformen Prädiktorqualität ausgegangen werden. Diese Befunde stehen einerseits in Einklang mit den Befunden von BOYCHUK (1991), LAMB et al. (1997) und ORBACH & LAMB (1999), in deren Studien dieses Kriterium sich als positiv bezüglich der Diskriminationsfähigkeit von wahren und falschen Aussagen erwies. Demgegenüber stehen jedoch eine Reihe konträrer Untersuchungsergebnisse, nach denen sich eine inadäquate Reliabilität dieses Realkennzeichens zeigte (vgl. STELLER et al., 1992; ANSON et al., 1993; LAMERS-WINKELMAN & BUFFING, 1996; HOROWITZ et al., 1997). Die genannten Befunde beziehen sich auf Untersuchungen mit in erster Linie normalbegabten Kindern als Versuchspersonen. Die Ergebnisse der vorliegenden Arbeit legen den Schluss nahe, dass es auch lernbehinderten Jugendlichen im Rahmen der Schilderung selbst erlebter Ereignisse zum einen leichter fällt bzw. dies natürlicher und nachvollziehbarer und zum anderen hinsichtlich ihrer kognitiven Leistungsfähigkeit einfacher ist, Nachbesserungen und Ergänzungen in Aussagen zu äußern als dies in erfundenen Aussagen konstruieren zu müssen (oder zu können).

Die post-hoc-Analyse der Veränderung über die beiden Zeitpunkte hinweg kommt zu dem Ergebnis, dass zwar für die wahren Geschichten eine statistisch bedeutsame Veränderung von T1 nach T2 zu verzeichnen ist, jedoch nicht für die erfundenen.

In fast alle referierten Studien, wie auch in der vorliegenden Arbeit, ergab sich im Ergebnis für das Realkennzeichen '*Zugeben von Erinnerungslücken*', dass dies zur Unterscheidung von erlebnisbasierten und erfundenen Schilderungen nur wenig beitragen kann. Eine Ausnahme stellt hier lediglich die Untersuchung von LAMERS-WINKELMAN & BUFFING (1996)

dar, in der für die 9-11jährigen Kinder recht hohe prozentuale Auftretenswerte festgestellt werden konnten. In der vorliegenden Untersuchung fanden sich zum einen weder deutliche Unterschiede zu den beiden Befragungszeitpunkten zwischen wahren und erfundenen Geschichten, zum anderen konnten weder für wahre noch für erfundene Geschichten signifikante Veränderungen zwischen den beiden Befragungszeitpunkten festgestellt werden. Trotz eines zum Teil zu verzeichnenden (nicht signifikanten) Anstiegs zum zweiten Befragungszeitpunkt könnte man für beide Realkennzeichen ('Spontanen Verbesserung der eigenen Aussage' und 'Zugeben von Erinnerungslücken') auch aufgrund ihres generell geringen Auftretens (vor allem bei 'Zugeben von Erinnerungslücken', deren gesamte Spanne im Rating nur von 0 bis 1 reichte) vermuten, dass hier die Autorität der Interviewerperson für die Jugendlichen eine entscheidende Rolle gespielt haben könnte. Für sie ist es ein besonderes und für alle erstmals erlebtes Ereignis, wenn ein Erwachsener, noch dazu jemand von der Universität, mit ihnen gewisse, wichtige Untersuchungen und Befragungen durchführt bzw. ihr Erinnerungsvermögen testen will bzw. sie persönliche, autobiographische Erlebnisse schildern sollen. Eventuell könnte auch die Instruktion bzw. die zusätzlich im Raum stehende Aufgabe „dem Interviewer eine erfundene Geschichte möglichst glaubhaft zu erzählen“ im Sinne einerseits einer spannenden und auch anspruchsvollen Aufgabe so auf die Jugendlichen gewirkt haben könnte bzw. andererseits diese eine solch hohe intrinsische Motivation entwickelt haben könnten, sich vor dem Interviewer oder anderen Personen keine Blöße geben zu wollen. Es wäre weiterhin zu prüfen, ob und inwieweit sich die Jugendlichen überhaupt veranlasst sahen Erinnerungslücken zuzugeben. Zwar wurden sie vom Interviewer darauf hingewiesen, dass es „nicht schlimm ist, wenn man mal etwas nicht mehr so genau weiß“. Jedoch bleibt fraglich, inwieweit sie dann solche Eingeständnisse verbal trafen. Natürlich gab es hinsichtlich der Motivation und Einsatzbereitschaft der Jugendlichen auch Unterschiede, jedoch konnte insgesamt von einer hohen Bereitschaft ausgegangen werden, nicht nur mitzumachen, sondern dem Interviewer gegenüber möglichst souverän aufzutreten dergestalt, dass dieser es schwer hatte, die beiden Geschichten hinsichtlich ihres Wahrheitsgehalts zu unterscheiden. Die Bandbreite der Qualität und Struktur der Versuche war so unterschiedlich und groß wie auch das Phänomen der Lernbehinderung selbst nur sein kann. Darauf wird im Kapitel zur Diskussion der Einschätzung der Glaubhaftigkeit noch näher einzugehen sein. Letztlich bleibt es nach GREUEL (2001) auch in weiteren Untersuchungen fraglich, ob sich die motivationale Situation eines falsch aussagenden Zeugen vor Gericht überhaupt experimentell induzieren lässt. Demnach dürfte sich das Befolgen einer Instruktion, im Rahmen einer Wettbewerbssituation möglichst gut und überzeugend zu „lügen“ aus psychologischer Sicht doch erheblich von der Motivation eines intentional falsch aussagenden Zeugen unterscheiden.

Zwar betrug die Spannweite der Merkmalsausprägung im Rating des Qualitätsmerkmals *‘Einwände gegen die Richtigkeit der eigenen Aussage’* nur etwa 0,5 Punkte, jedoch konnte zu beiden Befragungszeitpunkten ein trotzdem deutlicher Unterschied zwischen erlebnisbasierten Aussagen und Aussagen ohne Erlebnishintergrund gefunden werden, in letzteren also deutlich weniger derartige Einwände zu finden waren. Dies ist umso erstaunlicher bzw. bemerkenswerter, da die Jugendlichen in der ursprünglichen Instruktion und Anleitung durch Frau Dr. Loohs vermittelt bekamen, möglichst glaubwürdig beide Geschichten, eine tatsächlich erlebte und eine erfundene, zu erzählen und daher die Jugendlichen keine Veranlassung sahen, selbst an der Richtigkeit ihrer Aussagen zu zweifeln, da dies ja kontraproduktiv zum „Arbeitsauftrag“ wäre. Somit stellt sich zum einen erneut die Frage, inwieweit motivationsbezogene Inhalte in Simulationsstudien zu übertragen sind, so dass die Motive denen vor Gericht nahe kommen, zum anderen spricht dies für den Befund, dass gerade lernbehinderte Personen oder solche mit anderen intellektuellen Einschränkungen wohl nicht in der Lage sind, dies zu erkennen, zu berücksichtigen und ihre Aussage entsprechend zu manipulieren. Erneut konnten weder für wahre noch für erfundene Geschichten statistisch erkennbare oder relevante Veränderungen vom ersten zum zweiten Befragungszeitpunkt festgestellt werden. auch hier scheint der zeitliche Aspekt in den Aussagen der Jugendlichen nicht so sehr von Bedeutung zu sein.

Auch hinsichtlich des Aussagemerkmals *‘Selbstbelastungen’* kann man sich die Frage stellen, ob die motivationale Situation eines intentional falsch aussagenden Zeugen vor Gericht überhaupt in einer experimentellen Studie nachgestellt werden kann (vgl. GREUEL, 2001). Den Jugendlichen wurde in dieser Studie die Instruktion gegeben, zu versuchen, eine erfundene Geschichte und eine wirklich erlebte Geschichte so gut zu erzählen, dass der Interviewer Probleme bekommen sollte, den Wahrheitsgehalt beider zu erfassen bzw. voneinander zu unterscheiden.

Dementsprechend ließen sich zu keinem der beiden Befragungszeitpunkte bedeutsame Unterschiede zwischen den beiden Geschichten nachweisen.

Betrachtet man post-hoc die Unterschiede bezüglich der Veränderung der Merkmalsausprägung über die Zeit hinweg, so konnte hierbei für die wahren Geschichten eine signifikante, hypothesenkonforme Verringerung der Merkmalsausprägung beobachtet werden, während für die erfundenen Geschichten absolut keine Veränderung zu verzeichnen war.

Fast analog dazu verhalten sich die Befunde bezüglich des Kriteriums der *‘Inschutznahme des Täters’*. Mit Ausnahme weniger Studien (ANSON et al., 1993; BERGER, 2002) scheint dieses Merkmal zumindest in Simulationsstudien dieser Art nicht besonders trennscharf zu sein. Aufgrund der wirklich sehr geringen Auftretenshäufigkeit und Qualität des Merkmals in

beiden Geschichten und zu beiden Zeitpunkten (unter 0,3) ergaben sich weder Unterschiede zwischen den Aussagen noch Veränderungen zwischen den Befragungszeitpunkten.

Somit besteht die Möglichkeit, dass zum einen die Inhalte der Geschichten für die Jugendlichen keinen Ansatz boten einen Täter oder Beteiligten - soweit vorhanden - zu entlasten oder nicht weiter zu beschuldigen. Daher stellt sich die Frage, inwiefern die Themen der Geschichten nicht dafür verantwortlich sind, dass keine Täterentlastungen gefunden wurden. Bei der Auswahl der Themen sollte daher in weiteren Untersuchungen darauf geachtet werden, dass jedes Realkennzeichen überhaupt in Frage kommt und nicht von vornherein aufgrund der Inhalte ausgeschlossen wird.

‘Deliktsspezifische Aussageelemente oder Inhalte’ sind wohl wie kaum ein anderes Realkennzeichen von dem jeweiligen Vorfall, dessen emotionaler oder spezifischer Thematik sowie dem fachspezifischen Kenntnisstand des Beurteilers abhängig. Zwar betont STELLER (1999), dass in den meisten Simulationsstudien, in denen regelmäßig eine andere inhaltliche Thematik gegeben ist, dass das Auftreten von Merkmalen, die einen inhaltlichen Bezug zur Missbrauchsthematik haben – wie das bei den deliktsspezifischen Aussageelementen ja häufig der Fall ist bzw. die ursprünglich auf diesem Hintergrund konzipiert wurden (Anm. des Verfassers) – auch in erlebnisbasierten Schilderungen wenig wahrscheinlich ist. Demgegenüber ist nach ARNTZEN (1993) und Bender & Nack (1981, zit. nach GREUEL et al., 1998) das Merkmal *‘Deliktsspezifität’* nicht nur im Bereich der Sexualdelikte, sondern für den Gesamtbereich strafrechtlich relevanter Handlungen anwendbar.

Trotz der wiederum sehr geringen Ausprägungen des Qualitätsmerkmals in dieser Untersuchung, die sich in erster Linie aufgrund der - trotz der schon mehrfach genannten psychologisch bedeutsamen und in den ausgewählten Themen berücksichtigten Grundvariablen (nach FINKELHOR & BROWNE, 1986; STELLER et al., 1992) – Themenwahl erklären lassen, fanden sich auch in dieser Untersuchung mit kognitiv eingeschränkten Jugendlichen zu beiden Befragungszeitpunkten in erlebnisfundierten Aussagen deutlich mehr deliktsspezifische Aussageelemente als in nicht erlebnisfundierten. Somit kommt auch in Simulationsstudien bei richtigem Verständnis des Merkmals und entsprechender Anwendung auch auf nicht sexuelle Inhalte diesem eine gute Diskriminationsfähigkeit zu.

Eine deutliche Veränderung diesbezüglich vom ersten zum zweiten Befragungszeitpunkt konnte wie bei vielen Aussagemerkmalen in dieser Untersuchung nicht konstatiert werden.

Die Ergebnisse bezüglich des Gruppeneffekts in vorangegangenen Studien erweisen sich als recht unterschiedlich. Als positiv im Sinne einer Trennschärfe fand sich dieses Merkmal in den Studien von BOYCHUK (1991), HOROWITZ, LAMB et al. (1997) und LAMERS-WINKELMAN & BUFFING (1996) sowie in dieser Untersuchung. Andererseits trennte dieses

Merkmal nicht zwischen den Versuchsgruppen in den Studien von STELLER et al. (1992), ANSON et al. (1993) ORBACH & LAMB (1999) sowie TYE et al. (1999).

Zusammenfassend kann festgestellt werden, dass sich in dieser Untersuchung vor allem die Qualitätsmerkmale 'Logische Konsistenz', 'Widerspruchsfreie, ungeordnete Darstellung', 'Quantitativer Detailreichtum', 'Kontextuelle Einbettung', 'Schilderung nonverbaler Aktionen', 'Schilderung nebensächlicher Details', 'Schilderung eigener psychischer oder physischer Vorgänge', 'Einwände gegen die Richtigkeit der eigenen Aussage' und 'Deliktspezifische Aussageelemente' als relativ stabile Kriterien zur Differenzierung zwischen erlebnisfundierte und nicht erlebnisfundierte Aussagen erwiesen haben, da diese zu beiden Befragungszeitpunkten signifikant zwischen den Aussagen trennen.

Kritischer zu betrachten bleiben diejenigen Kriterien, die entweder nur zum ersten Befragungszeitpunkt T1 zwischen wahren und falschen Aussagen trennen konnten ('spontane Verbesserungen der eigenen Aussage') oder nur zum zweiten Befragungszeitpunkt T2 ('Wiedergabe von Gesprächen', 'Schilderung von Komplikationen im Handlungsverlauf' sowie die 'Schilderung ausgefallener Details').

Betrachtet man die Befunde dieser sowie der anderen Untersuchungen im Einzelnen, so stellt man fest, dass selbst in erlebnisbasierte Aussagen kaum jemals alle Qualitätsmerkmale vorkommen. Dazu kommt nach KÖHNKEN (2001) auch noch folgende Problematik, nämlich dass in den unterschiedlichen Studien jeweils verschiedene Merkmale statistisch signifikante Differenzen zwischen wahren und erfundenen Aussagen aufweisen. Dies ist jedoch insofern nicht überraschend, da bei der Aussageanalyse im Allgemeinen auf inhaltliche Aspekte Bezug genommen wird, die sich natürlich von Fall zu Fall unterscheiden und demnach nicht unabhängig von dem jeweils berichteten Sachverhalt sein können. Deshalb ist eine Aussage pauschal nicht weniger glaubhaft, nur weil bestimmte Realkennzeichen nicht darin vorkommen. Von daher erfassen nach STELLER (1999) solche Bewertungen, die postulieren, dass diejenigen Merkmale am wichtigsten seien, die in den meisten Untersuchungen zwischen wahren und falschen Schilderungen unterscheiden, die eigentliche Bedeutung und Aussagekraft der Realkennzeichen nur unzureichend, da sie die unterschiedlichen inhaltlichen Validitäten der einzelnen Merkmale nicht genügend berücksichtigen.

Demzufolge erscheint der von manchen Forschern gemachte Vorschlag wenig sinnvoll, diejenigen Kriterien aus den Merkmalslisten zu streichen, die in bestimmten einzelnen Studien aufgrund verschiedener individueller Faktoren sich als nicht genügend trennscharf erwiesen haben. In anderen Situationen, Konstellationen oder mit anderen thematischen Inhalten können diese vermeintlich unbedeutameren Merkmale durchaus wertvolle und valide Diskriminatoren darstellen.

Da dies neben ERMANN (2005, in Vorbereitung) die bisher einzig bekannte Simulations- und Validierungsstudie bei lernbehinderten Jugendlichen ist, gilt es in weiteren Forschungsarbeiten zu prüfen, ob sich die gefundenen diskriminativen Merkmale als allgemeingültige Diskriminatoren bestätigen können oder ob die diesbezüglichen Ergebnisse der vorliegenden Arbeit auf verschiedene andere Faktoren wie Stimulusmaterial, Interviewerverhalten oder Fehler beim Rating etc. zurückführen lassen.

Gerade im Zusammenhang mit der diesbezüglich noch sehr wenig untersuchten speziellen Population der Lernbehinderten eröffnen sich neue und breite Forschungsfelder. Intensive Forschungsarbeit ist aufgrund der großen, individuell recht unterschiedlichen und heterogenen Gruppe der Menschen mit Intelligenzminderung notwendig.

5.2.2 Diskussion zu den Befunden über alle Aussagemerkmale hinweg und zu den Merkmalskategorien

Wie bereits dargestellt kann nach dem Aggregationsprinzip ebenfalls untersucht werden, ob alle in die statistische Analyse einbezogenen Qualitätsmerkmale auch in Kombination bzw. als Gesamtinstrument in der Lage sind, gemäß der 'Undeutsch-Hypothese' zwischen erlebnisbegründeten Aussagen und Aussagen ohne Erlebnishintergrund zu trennen. Dazu wurden die Mittelwertsausprägungen aller Aussagemerkmale jeweils zu einem Summenscore zusammengefasst und zunächst varianzanalytisch untersucht, wobei sich jeweils deutliche Unterschiede hinsichtlich der Gruppenzugehörigkeit der Aussagen (wahr oder erfunden) wie auch des Befragungszeitpunktes ergaben. Somit sind in erlebnisfundierten Aussagen die Realkennzeichen deutlich stärker ausgeprägt sind als in den Aussagen ohne Erlebnishintergrund. Dies gilt sowohl über beide Befragungszeitpunkte hinweg betrachtet als auch für beide Zeitpunkte einzeln. Dies bestätigt die insgesamt doch eindeutigen Befunde aus den einzelnen Realkennzeichen, wodurch die Analyse der aussageimmanenten Qualitätsmerkmale als Teil der integrativen Krieriologie auch bei der Personengruppe der lernbehinderten Jugendlichen bei geschulter und wissensbasierter Anwendung als geeignetes Instrument im Rahmen der forensischen Glaubhaftigkeitsbegutachtung angesehen werden kann und muss. Die Ergebnisse dieser Untersuchung sind im Gesamten deckungsgleich mit vielen vorangegangenen Feld- und Simulationsstudien vor allem bei Kindern, die alle die Validität der 'Undeutsch-Hypothese' bzw. des Kriterienkataloges von Steller & Köhnken (1989) weitestgehend bestätigten. Demnach weisen auch in dieser Untersuchung diejenigen Aussagen mit einem entsprechenden Erlebnishintergrund jeweils höhere Ausprägungen in den Qualitätsmerkmalen auf als die erfundenen Aussagen.

An dieser Stelle sei auf die Anmerkung von DAHLE (1997) hingewiesen, nachdem bei dieser Methode der Summenscorebildung, die einzelnen Werte (Mittelwertsausprägungen) natürlich ungewichtet in die statistische Analyse eingehen. Dies ist gemäß des Aggregationsprinzips zwar möglich, jedoch auch kritisch zu betrachten, da wie ausführlich dargestellt zum einen die Merkmalsanalyse nicht als reines Checklistenverfahren fehlinterpretiert werden darf, zum anderen die einzelnen Aussagemerkmale wie auch die verschiedenen Kategorien gerade in der forensischen Gutachtenpraxis durchaus qualitativ unterschiedlich je nach im Raum stehendem Tatvorwurf oder auch individuellen Kompetenzen, Bedürfnissen und Motiv(ations)lagen der Zeugen bewertet werden (müssen). Jedoch sind im aussagepsychologischen Kontext in erster Linie die 'Allgemeinen Merkmale' sowie 'Deliktspezifische Inhalte' immer von größter Bedeutung und Aussagekraft. Die übrigen Merkmale müssen dabei wie gesagt immer von individuellen Faktoren und Bedingungen abhängig gemacht werden. Dabei wurde in der vorliegenden Arbeit bereits desöfteren auf die Schwierigkeit im Zusammenhang mit der Übertragbarkeit vor allem der motivationsbezogenen Inhalte von Simulationsstudien auf tatsächliche (motivationale) Situationen vor Gericht hingewiesen.

Ebenso deutlich wie die Befunde der Differenzierbarkeit des Wahrheitsgehalts der Aussagen generell, kann auch das Ergebnis bezüglich der Veränderung über die beiden Befragungszeitpunkte hinweg gesehen werden. Auch hier ergibt die post-hoc-Analyse für die wahren ebenso wie für die erfundenen Geschichten eine deutliche Veränderung von T1 nach T2 in Form eines Abfalls der Merkmalsausprägungen. Zwar konnte im Rahmen der Interviewsituation zum zweiten Befragungszeitpunkt von bestimmten Lerneffekten in Form von Erwartungen und Kenntnis der Situation und Person des Interviewers, der Räumlichkeiten und Abläufe der Befragung ausgegangen werden, jedoch waren die Jugendlichen offenkundig nicht in der Lage, daraus sichtbaren Nutzen zu ziehen. Zwar wurde erwartet, dass sich bei den erlebnisbasierten Aussagen nur kleine im normalen Rahmen des Vergessens übliche geringere Merkmalsausprägungen finden, jedoch keine wesentlichen, da angenommen werden kann, dass diese aus dem autobiographischen Gedächtnis abgerufen werden. Es fanden sich allerdings wie auch bei den nicht erlebnisfundierte Geschichten in der Summe deutliche Abnahmen von T1 nach T2. Somit kann daraus gefolgert werden, dass die Jugendlichen bei der Schilderung beider Arten von Geschichten Vergessensprozessen ausgesetzt waren, was bezüglich der erfundenen Geschichten weniger verwundert. Es kann davon ausgegangen werden, dass Jugendliche mit einer Intelligenzminderung nicht in der Lage sind, weder selbst erlebte noch erfundene Inhalte über einen Zeitraum von sechs Wochen so im Gedächtnis zu speichern und sie auch erneut in ähnlicher Qualität wieder reproduzieren können. Es ist anzunehmen, dass nicht nur der zeitliche Abstand von sechs Wochen von T1 zu T2 dazu führt, dass gerade bei den nicht-erlebnisbasierten Aussagen Vergessensprozesse eingesetzt haben und der Abruf aus dem semantischen Gedächtnis im Vergleich zum Abruf

aus dem autobiographischen Gedächtnis bei den erlebnisbasierten Geschichten erschwert ist. Trotzdem ist aber anzunehmen, dass die nicht-erlebnisbasierten Aussagen von den Jugendlichen weniger häufig erneut reflektiert wurden als die wirklich erlebten, da das Erlebnis mit realem persönlichen Hintergrund ihnen näher zu sein schien.

Lediglich in 6 der 17 in die statistische Analyse einbezogenen Aussagemerkmale konnte kein signifikanter Gruppenunterschied festgestellt werden. Dazu muss jedoch angemerkt werden, dass dies zum Teil an der Tatsache des Vorliegens einer Simulationsstudie lag, zum Teil an entsprechenden Instruktionen der Versuchspersonen bzw. auch an der Wahl und den Inhalten der Themen (Geschichten).

Eine wechselseitige Abhängigkeit der beiden Faktoren konnte auch über alle Kriterien nicht festgestellt werden.

So kann zusammenfassend festgehalten werden, dass trotzdem manche Realkennzeichen nicht zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen trennen konnten, die Methode der kriterienorientierten Aussageanalyse als Gesamtkonzept auch bei Personen mit kognitiven Defiziten als gut anwendbar, valide und reliabel betrachtet werden kann.

Fasst man die ersten der Realkennzeichen zu der Kategorie *'Allgemeine Merkmale'* zusammen, so lassen sich analog zu obiger Diskussion zum einen feststellen, dass diese, denen in der aussagepsychologischen Literatur große Bedeutung zugewiesen wird, da sich diese auf die Aussage als Gesamtprodukt und deren globales Erscheinungsbild beziehen, auch in dieser Untersuchung (höchstsignifikant) zwischen wahren und falschen Aussagen trennen können – und zwar zu beiden Befragungszeitpunkten. Dies entspricht ebenso der *'Undeutsch-Hypothese'*, den übrigen Ergebnissen verschiedener Feld- und Simulationsstudien sowie den Befunden aus der Gedächtnispsychologie und Informationsverarbeitung. Im Vergleich und im Hinblick auf andere Untersuchungen und deren Design oder Versuchsmaterial kann – eben unabhängig vom Versuchs- oder Stimulusmaterial – davon ausgegangen werden, dass diese Kriterien natürlicherweise und selbsterklärend in allen erlebnisbasierten Aussagen vorkommen können, was nicht für alle Realkennzeichen gilt. Analog zu den Einzel-Befunden der ersten drei Merkmale finden sich auch bei Betrachtung der Kategorie der *'Allgemeinen Merkmale'* keine Veränderungen über die Zeit hinweg, zwischen den beiden Befragungszeitpunkten und zwar weder für wahre noch für erfundene Geschichten. Dies spricht für eher stabile charakteristische Eigenschaften der ersten drei Qualitätsmerkmale. Diese Merkmale kommen in erlebnisbasierten Aussagen (häufiger) wie auch in nicht erlebnisbasierten Aussagen (deutlich weniger häufig) vor, scheinen jedoch in dieser Untersuchung (Stichprobe) weitgehend unabhängig vom zeitlichen Einfluss oder sind Veränderungen von einem gegenüber einem nachfolgenden Befragungstermin weniger unterworfen. Möglicherweise kann daraus gefolgert werden, dass so gut (in wahren Geschichten) bzw.

schlecht (in erfundenen) die qualitative Struktur von Aussagen (hinsichtlich der bedeutsamen Kriterien 'Logische Konsistenz', Widerspruchsfreie, ungeordnete Darstellung' und 'Quantitativer Detailreichtum') lernbehinderter Jugendlicher einmal ist, so bleibt sie auch – zumindest über einen gewissen Zeitraum von sechs Wochen. Ab welchem zeitlichen Abstand dieser Effekt eventuell kippt, kann Aufgabe und Gegenstand zukünftiger längsschnittlicher Forschungsarbeiten sein.

Trotzdem sollte nach Möglichkeit immer versucht werden, im Falle mehrerer notwendiger Befragungen, diese nicht unnötig zu verzögern, da sich selbstverständlich die Zeit ab einem gewissen Zeitpunkt als negativer Einflussfaktor auf die Gedächtnisleistung auswirkt.

Offensichtlich stehen Effekte der Reminiszenz und Hypermnese in direktem Zusammenhang mit natürlichen Vergessenprozessen, wirken sich jedoch nur indirekt aus, indem sich bis zu einem gewissen Zeitpunkt, neu erinnerte Gedächtnisinhalte, Vergessenprozesse und andere Dekodierungsvorgänge sowie rekonstruktive Vorgänge in einem noch genauer zu erforschenden Verhältnis die Waage halten und sich soweit in relativ stabilen Erinnerungsleistungen manifestieren.

Die Befunde zu den Kategorien '*Spezielle Inhalte*' und '*Inhaltliche Besonderheiten*' können simultan betrachtet werden. Zusammengenommen scheinen die Kriterien dieser beiden Kategorien jeweils sehr deutlich sowohl zwischen den Aussagen wie auch zwischen den Befragungszeitpunkten trennen zu können. Post-hoc-Analysen ergaben zum einen zu beiden Befragungszeitpunkten deutlich geringere Mittelwertsausprägungen in der Kategorie für Geschichten ohne Erlebnisbasis gegenüber solchen, die tatsächlich erlebt wurden. Zum anderen scheinen diese beiden Kategorien im Gegensatz zu den 'Allgemeinen Merkmalen', den 'Motivationsbezogenen Inhalten' wie auch den 'Deliktspezifischen Inhalten' durchaus dem zeitlichen Einfluss eines oder mehrerer späterer Befragungszeitpunkte zu unterliegen, da sich aufgrund von post-hoc-Analysen für beide Kategorien in wahren wie in erfundenen Geschichten zu einem zweiten Befragungszeitpunkt T2 in der Summe deutlich geringere Merkmalsausprägungen fanden als zum ersten Befragungszeitpunkt T1.

Der schon mehrfach erwähnten Merkmalskategorie der '*Motivationsbezogenen Inhalte*' und deren besonderer Stellung und Problematik im Rahmen von Simulationsstudien kommt auch bei simultaner (kumulierter) Betrachtung der einzelnen Realkennzeichen dieser Gruppe eine interessante Bedeutung zu. Nach GREUEL et al. (1998) kommt den 'Motivationsbezogenen Inhalten' auch eine geringere Belegkraft für einen möglichen Erlebnisbezug zu, da sich diese nicht unmittelbar durch kognitive Gesetzmäßigkeiten begründen lassen (vgl. GREUEL et al., 1998). Die Argumentation für diese Merkmalskategorie basiert auf der Annahme, dass ein intentional falsch aussagender Zeuge in der Regel versuchen wird, seine Person vor

Gericht als besonders glaubwürdig darzustellen und gemäß seiner subjektiven Glaubwürdigkeitskonzepte sich besonders positiv zu präsentieren versucht.

So ist diese Merkmalskategorie in der vorliegenden Untersuchung nur zum ersten Befragungszeitpunkt in der Lage, zwischen wahren und erfundenen Geschichten zu diskriminieren. Zwar wurde auch in dieser Studie versucht, die Jugendlichen entsprechend zu motivieren und ihnen die Wichtigkeit ihrer „Zeugenaussage“ vor Augen zu halten, jedoch kann nicht ausgeschlossen werden, dass unkontrollierte Motive vorlagen. Auch im Rahmen der Befragungen konnte neben den erwähnten Lern- und Intervieweffekten und insgesamt guter Kooperationsbereitschaft bei einigen Jugendlichen vielleicht im Ansatz weniger Konzentration und „Anspannung“ in der nun schon bekannten Interviewsituation beobachtet werden. Da die Untersuchung der Glaubhaftigkeit von Zeugenaussagen nach der Untersuchung zur Erinnerungsleistung bzw. der Quellenidentifikation stattfand, hatten manche der Befragungen auf Seiten der Jugendlichen bereits eine Art schematisierten und weniger natürlich-spontanen Charakter oder Erzählstil, trotzdem natürlich von Seiten des Interviewers versucht wurde, dem anhand entsprechender Frage- und Interviewtechniken entgegenzusteuern.

So ist es auch wenig verwunderlich bzw. teilweise nachvollziehbar, dass sich weder für wahre noch für erfundene Geschichten eine deutliche Veränderung der motivationsbezogenen Inhalte vom ersten zum zweiten Befragungszeitpunkt feststellen ließ.

Die motivationsbezogenen Qualitätsmerkmale stellen somit für experimentelle Untersuchungen zur kriterienorientierten Aussageanalyse eine besondere Herausforderung dar. Dies kann und soll in weiteren Untersuchungen durchaus kritisch bedacht werden und auf den besonderen Charakter dieser Merkmalskategorie hinweisen. Speziell sollte dies für solche spezifischen Personenkreise wie etwa den der lernbehinderten Jugendlichen sorgfältig und weiter untersucht werden. Wenn sich motivationale Bedingungen bereits in simulierten Validierungsuntersuchungen als derart kritisch erweisen, impliziert dies doch wohl eine entsprechende Tragweite vor Gericht bzw. im Begutachtungskontext. Neben den auch kritisch zu beachtenden, insgesamt wohl aber eher positiv zu interpretierenden Aspekten der Vertrautheit mit der Person des Interviewers, den groben Abläufen von Befragungen und anderen bekannten Faktoren wären in diesem Zusammenhang beispielsweise zu diskutieren bzw. immer neu und aktuell zu überdenken: Kann auch bei weiteren Befragungen von einer positiven und angemessenen Motivation des Zeugen ausgegangen werden, im Sinne einer Kooperation und im Rahmen seiner möglichen Kompetenzen, eine ausführliche und möglichst detailgenaue Schilderung abgeben zu wollen? Ist dies nicht der Fall, muss überlegt werden, wie und ob gegebenenfalls ein neuer Anreiz gesetzt werden kann.

Auch wenn in verschiedenen Simulationsstudien dieser Merkmalskategorie keine allzu große Trennschärfe zugesprochen wurde und sich die Effekte der Ergebnisse der vorliegenden Arbeit ebenfalls im Rahmen halten, drängt sich doch der gerechtfertigte Verdacht auf, dass

gerade bei Jugendlichen und noch spezieller bei solchen mit intellektuellen Beeinträchtigungen diese Merkmalskategorie von besonderer Relevanz ist und in keinem Fall, wie von manchen Forschern vorgeschlagen (wie bereits erwähnt), aus der Liste der Merkmale zu entfernen. Stattdessen sollte der Focus vielmehr auf dem fundierten Ausbau methodisch gesicherter Studien mit speziellen, besonders relevanten Personengruppen in ökologisch validen Situationen und vergleichbaren thematischen Inhalten liegen. In zukünftigen experimentellen Forschungsarbeiten sollte den 'Motivationsbezogenen Inhalten' somit quasi die Chance eröffnet werden, ihr – nach Meinung des Verfassers – doch im Rahmen der forensischen Praxis und Realität großes und wichtiges Potential hinsichtlich kritischer Personengruppen darzulegen.

Zwar ist das Vorkommen '*deliktspezifischer Inhalte*' nach deren klassischem Verständnis in Simulationsstudien oftmals kaum bzw. in geringer Qualität (wie auch in dieser Untersuchung) zu erwarten, jedoch fanden sich dennoch zu beiden Befragungszeitpunkten immer noch deutliche Unterschiede in der Merkmalsausprägung von erlebnisbasierten Aussagen und Aussagen ohne Erlebnishintergrund. Eine zeitliche Veränderung zwischen den Befragungszeitpunkten ergab sich sowohl für erlebnis- wie für nicht erlebnisfundierte Aussagen nicht. Die genannten Befunde spiegeln sich in entsprechend geringen Effektstärken wieder. Eine diesbezüglich entsprechend reale Simulation oder Abbildung tatsächlicher, forensisch relevanter Geschehnisse erweist sich unter ethischen wie auch aus technischen Gründen des Untersuchungsdesigns schwer möglich und problematisch.

Zusammenfassend kann festgestellt werden, dass Simulationsstudien zwar durchaus eine geeignete Methode darstellen die Validität von Glaubhaftigkeitskriterien zu prüfen, jedoch müssen Simulationsstudien immer mit Zurückhaltung interpretiert werden. Eine „1:1“-Übertragung auf die Wirklichkeit und somit auf die gerichtliche Praxis ist nicht realistisch. Eine gewisse „Unschärfe“ der Abbildung der Wirklichkeit kann nicht vermieden werden.

5.2.3 Diskussion zu den Befunden der diskriminatorisch bedeutsamen Aussage-merkmale, den Klassifizierungsergebnissen und Glaubhaftigkeitseinschätzungen

Wie auch in dieser Untersuchung wurden ebenfalls in einigen, vor allem aktuelleren Studien (z.B. TYE et al., 1999; AKEHURST et al., 2001) in der weiteren statistischen Analyse Diskriminanzanalysen berechnet, in denen zum einen die diskriminatorisch bedeutsamen Real-kennzeichen nach ihrer Potenz geordnet erfasst wurden, zum anderen die jeweiligen korrek-

ten und falschen Klassifikationsraten im Sinne der Signalentdeckungstheorie vergleichend gegenübergestellt.

Betrachtet man dabei im Ergebnis z.B. die ersten acht bzw. neun Qualitätsmerkmale, die sich am besten zur Trennung zwischen wahren und falschen Schilderungen erwiesen haben, ergeben sich im Vergleich mit anderen Studien zum Teil interessante Unterschiede und Abweichungen, aber auch gewisse konkrete und tendenziell ähnliche Ergebnisse. Verschiedene Befunde sind dabei besonders auffällig und bemerkenswert:

An erster Stelle gilt es zu bemerken, dass zu beiden Befragungszeitpunkten die Kriterien der ersten Kategorie, der 'Allgemeinen Merkmale' sich als besonders trennscharf und diskriminatorisch bedeutsam herausgestellt haben. Diese Aussagemerkmale beziehen sich ja grundsätzlich auf die Gesamtheit der Aussage und deren logischem Aufbau und Nachvollziehbarkeit, stellen also das Grundgerüst jeder glaubhaften Aussage und sind deshalb von so großer Bedeutung. Sind im forensischen Kontext diese Merkmale qualitativ nur unzureichend ausgeprägt, ist die Glaubhaftigkeit der entsprechenden Aussage von vornherein grundsätzlich in Zweifel zu ziehen. In dieser Untersuchung nehmen diese ersten drei Qualitätsmerkmale (Logische Konsistenz, widerspruchsfreie, ungeordnete Darstellung und quantitativer Detailreichtum) zum ersten Befragungszeitpunkt sogar die ersten drei Plätze hinsichtlich der diskriminatorisch bedeutsamen Merkmale ein, zum zweiten Zeitpunkt die Plätze 1, 4 und 8. Dass die Menge der berichteten Einzelheiten zu T2 in der Bedeutung ein wenig abfällt, verwundert nicht und ist mit natürlichen Vergessensprozessen zu erklären. Offenkundig sind diese elementaren Aussagemerkmale für Jugendliche mit kognitiven Defiziten bei der Schilderung ihrer Aussagen von zentraler Bedeutung und erhalten daraus ihren hohen diagnostisch-diskriminatorischen Wert. Im Falle erlebnisfundierter Ereignisse kommen die genannten Merkmale in entsprechenden Schilderungen naturgemäß in hoher Qualität vor, im Falle von Schilderungen von Ereignissen ohne Erlebnishintergrund fällt es diesem Personenkreis aufgrund ihrer eingeschränkten intellektuellen Kompetenzen und Planungskapazitäten sehr schwer, logisch konsistente, widerspruchsfreie und detailreiche Aussagen zu produzieren bzw. glaubhaft darzustellen.

Weiterhin kommen auch zu beiden Befragungszeitpunkten besonders Merkmalen aus dem Bereich der 'Speziellen Inhalte' eine große Bedeutung zu, hier vor allem den Merkmalen 'Kontextuelle Einbettung' und 'Schilderung nonverbaler Aktionen'. Besonders interessant ist hier die Tatsache, dass es für lernbehinderte Jugendliche offensichtlich zu einem späteren zweiten Zeitpunkt einer Aussage sehr schwierig ist, eine zeitlich-räumliche Einordnung eines nicht selbst erlebten Geschehens in eine entsprechende erfundene Geschichte plausibel einzubauen. Hier finden sich große Unterschiede im Vergleich zu Schilderungen erlebnisfundierter Ereignisse. Dementsprechend kommt dem Merkmal der 'Kontextuellen Einbettung' nach der 'Logischen Konsistenz' die wichtigste diskriminative Funktion zu. Dies deckt sich

auch mit einigen anderen Befunden (TYE et al., 1999; BERGER, 2002), wonach diesem Kriterium sogar das bedeutsamste Gewicht zukommt. Für einen Zeugen ohne wirklichen Erlebnishintergrund scheint es tatsächlich erheblich schwerer zu sein, die situative und zeitliche Anordnung und den Ablauf der Geschehnisse schlüssig in einen Zusammenhang bringen zu können und dies plausibel darzustellen, selbst wenn er - wie in Simulationsstudien ja oft der Fall - gewisse Hinweisreize gegeben bekommt.

Ebenfalls interessant ist der Befund, dass aus dem Bereich der 'Inhaltlichen Besonderheiten' nur die 'Schilderung eigener psychischer oder physischer Vorgänge' eine - jedoch offensichtlich relevante - Rolle spielt (mit Ausnahme der 'Schilderung ausgefallener Details' zu T2). Wie bereits im Rahmen der Diskussion zu den einzelnen Realkennzeichen erwähnt, scheinen jugendliche Lernbehinderte zum einen bereits bei der Enkodierung eines Ereignisses wie auch beim Abruf vorrangig auf ihre eigenen emotionalen Vorgänge und Reaktionen fokussiert zu sein und eher weniger auf externale Vorgänge oder Prozesse bei anderen Beteiligten oder Tätern. Zum anderen fällt es ihnen somit eventuell auch schwerer, sowohl eigene psychische Empfindungen in erfundene Aussagen als auch psychische oder physische Vorgänge bei anderen Personen in wahre wie in erfundene Geschichten zu transferieren bzw. solche zu konstruieren. So könnten die großen Unterschiede zwischen erlebnisfundierten und nicht erlebnisfundierten Aussagen zu beiden Befragungszeitpunkten erklärbar sein.

Die mittlere diskriminatorische Funktion der 'Schilderung ausgefallener Details' zum zweiten Befragungszeitpunkt steht möglicherweise mit folgender Überlegung in Zusammenhang. Da über die Zeit hinweg die Gesamtmenge an erinnerten Details natürlicherweise abnimmt, sehen Lernbehinderte wie wohl auch normal Begabte sich eher dazu veranlasst im Rahmen dieses Vergessens, sich an außergewöhnlichere Details eines tatsächlich erlebten Geschehens oder solche mit einer geringeren Auftretenswahrscheinlichkeit zu erinnern oder sich erinnern zu können. In konstruierten Aussagen übersteigt dies meist schon den Planungshorizont intentional falsch aussagender Menschen. Entsprechend schwieriger gestaltet sich dies bei Lernbehinderten aufgrund ihrer kognitiven Defizite.

Aus dem Bereich der vieldiskutierten 'Motivationsbezogenen Inhalte' zeichnen die beiden Merkmale 'Spontane Verbesserungen der eigenen Aussage' zum ersten Befragungszeitpunkt sowie 'Einwände gegen die Richtigkeit der eigenen Aussage' zum zweiten Befragungszeitpunkt ein natürliches und zu erwartendes Bild. Traten zu T1 in den wahren Geschichten noch vermehrt spontane Verbesserungen der eigenen Aussage gegenüber den erfundenen auf, konnten diese zu T2 nicht mehr so deutlich differenzieren. Jedoch fanden sich (stattdessen?) vor allem mehr Einwände gegen die Richtigkeit der eigenen Aussage in erlebnisbasierten Aussagen gegenüber nicht erlebnisbasierten. So wäre es möglich, dass die Versuchspersonen zu einem zweiten Befragungszeitpunkt bei der Schilderung und

Rekonstruktion der wahren Geschichte versuchten, sich an ihre erste Darstellung zu erinnern und dies dann in vielen Fällen zu Unsicherheiten führte.

Diese Untersuchung zeigt also, dass weiterhin davon ausgegangen werden muss, dass auch die 'Motivationsbezogenen Inhalte' in bestimmten Untersuchungssettings und -bedingungen recht gut zur Diskrimination von wahren und falschen Aussagen beitragen können. Dies gilt in dieser Arbeit nicht nur für den ersten Befragungszeitpunkt, sondern auch für den zweiten, was z.B. wiederum im Gegensatz zu den entsprechenden Befunden von AKEHURST et al. (2001) steht. Dort kam dieser Kategorie von Realkennzeichen kein starker Diskriminationseffekt zu. Dies spricht dafür, in weiteren Studien vermehrt den Fokus auf 'Motivationsbezogene Inhalte' zu setzen und diese gesondert systematisch zu untersuchen, speziell im Hinblick auf den forensisch-praxisrelevanten und vieldiskutierten Aspekt einer zweiten, wiederholten Befragung. Was die Befunde zu den 'Motivationsbezogenen Inhalten' angeht, so unterstützt die vorliegende Arbeit den von KÖHNKEN (2001) angeführten und schon zitierten Kritikpunkt, dass es unsinnig sei, z.B. der Auffassung von GREUEL et al. (1998) oder anderen Autoren zu folgen und bestimmten Merkmalen, etwa den 'Motivationsbezogenen Inhalten' eine geringere Belegkraft für den möglichen Erlebnisbezug einer Aussage zuzuschreiben als anderen. Dazu müssten erst genügend aussagekräftige empirische Belege vorliegen.

Weiterhin sei auf die verschiedenen möglichen Einflussgrößen auf den Wahrnehmungs- und Wiedergabeprozess von Zeugen hingewiesen. Als einige Aspekte sind hier die Erwartungshaltung des Zeugen, dessen Aufmerksamkeitsverteilung oder sein emotionales Befinden zu nennen, die sich eventuell negativ auf die Aussagequalität auswirken können. Im Rahmen dessen weiß man aus Befunden der Gedächtnispsychologie etwa, dass die menschliche Erinnerung stimmungs- und kontextabhängig ist, d.h. dass der Zeuge sich zum Zeitpunkt der Befragung im gleichen emotionalen Zustand befinden sollte wie zum Zeitpunkt der Informationsspeicherung (Kongruenz bzw. Inkongruenz von Kontexten). Gibt es hier Differenzen, so müssen Qualitätseinbußen in der Zeugenaussage einkalkuliert werden.

Bezüglich der Klassifizierungsquoten der Diskriminanzanalysefunktion, lässt sich sagen, dass im Mittel über beide Befragungszeitpunkte 96 % aller Aussagen richtig klassifiziert worden sind und demnach 4 % nicht korrekt. Betrachtet man dieses Ergebnis detaillierter, d.h. z.B. nach falschen und wahren Aussagen, so stellt man fest, dass ebenfalls über beide Zeitpunkte gesehen die erfundenen Schilderungen zu 98 % als solche identifiziert wurden, die erlebnisbasierten zu 94 % korrekt als solche. Vergleicht man diese Quoten mit den Ergebnissen anderer Studien, so kann dies als sehr gutes Ergebnis gewertet werden, das sich auch in etwa mit dem anderer Untersuchungen deckt. So ergeben sich etwa bei TYE et al. (1999) für die falschen Aussagen zwar eine 100%ige Aufdeckungsquote, für die wahren jedoch nur eine 75%ige. Im Schnitt wurden in dieser Untersuchung demnach trotzdem 96 %

aller Aussagen richtig klassifiziert. Bei AKEKURST et al. (2001) ergeben sich getrennt für die Altersgruppe der Kinder, dessen Vergleich hier relevant ist, allerdings nur 71,4 % korrekte Klassifizierungen aller Aussagen über die Zeit.

Allerdings stellt sich in dieser Diskussion die Frage, ob ein solcher gemittelter Vergleich überhaupt sinnvoll erscheint, da in dieser Untersuchung ja zwei Befragungszeitpunkte vorlagen, die durch ein relativ langes Zeitintervall voneinander entfernt waren. Jedoch ergeben sich bei separater Analyse der einzelnen Zeitpunkte auch sehr interessante Feststellungen. Zum T1 ist die Diskriminanzfunktion noch in der Lage sehr gute 96 % aller Aussagen richtig zu klassifizieren, wobei diese Quote mit bei 96 % für die wahren Aussagen ('Hits') gegenüber ebenfalls 96 % für die falschen ('Correct Rejections') identisch ausfällt. Bei der Gesamtquote der korrekten Klassifizierungen ist zu T2 insgesamt kein Unterschied zu verzeichnen, jedoch ergibt sich hinsichtlich des Verhältnisses von 'Hits' und 'Correct Rejections' ein kleiner Unterschied. Das Verhältnis der 'Hits' (92 %) beträgt gegenüber den 'Correct Rejections' (100 %) nur acht Prozentpunkte weniger. Offensichtlich wird es zu einem zweiten Befragungszeitpunkt T2 im Vergleich zu T1 etwas leichter, falsche Aussagen als solche zu klassifizieren, jedoch bewegt sich die Quote auch wahre Aussagen als wahre zu klassifizieren mit nur vier Prozentpunkten weniger gegenüber einem ersten Befragungszeitpunkt auf einem ähnlich hohen Niveau.

Betrachtet man die eben dargestellten Klassifizierungsergebnisse im Vergleich zu den tatsächlichen Einschätzungen des Raters (= Verfasser der Arbeit), so ergibt sich zunächst über beide Befragungszeitpunkte hinweg gemittelt, dass die Ratereinschätzung (von korrekt identifizierten Aussagen, also 'hits' und 'correct rejections') mit 89 % zwar unter der Quote der Diskriminanzfunktion (96 %) liegt, jedoch nur gering. Analog zu obigen Diskussionsbefunden wurden von den erlebnisbasierten Aussagen 90 % korrekt als solche erkannt, von den erfundenen 88 % korrekt als Falschaussagen. Dabei scheint es über beide Zeitpunkte hinweg tendenziell zwar etwas einfacher zu sein, eine Aussage korrekt als glaubhaft zu diagnostizieren (45 %), als die Unglaubwürdigkeit einer Aussage korrekt festzustellen (44 %), jedoch besteht fast kein Unterschied. Insgesamt kann festgestellt werden, dass in dieser Untersuchung nur marginale Unterschiede zu beiden Befragungszeitpunkten bestehen, wahre wie auch erfundene Aussagen korrekt jeweils als solche zu identifizieren. Tendenziell steigen bzw. fallen die Quoten der 'hits' und 'correct rejections' bzw. der 'false alarms' 'misses' leicht an bzw. ab, allerdings scheint die Qualität der Aussagen (sowohl der erlebnisfundierten wie der nicht erlebnisfundierten) von lernbehinderten Jugendlichen über die beiden Befragungszeitpunkte relativ stabil zu sein, dergestalt, dass diese in etwa gleich gut voneinander unterschieden werden können. Möglicherweise verändern sich diese Quoten in einer dritten oder vierten Befragung stärker oder stehen in Zusammenhang mit dem Schwe-

regard der intellektuellen Beeinträchtigung der Versuchspersonen oder anderen persönlichen und/oder sozialen Faktoren. Weitere, vor allem längsschnittliche Untersuchungen, könnten hier konkrete, weiterführende empirische Belege für praxisrelevante Implikationen liefern. Dennoch sollten grundsätzlich aufeinander folgende Befragungen zumindest nicht unnötig hinausgezögert werden, um eventuell zu der Erkenntnis zu gelangen, ob gerade bei Menschen mit kognitiven Defiziten ein für Zeugenaussagen und deren Begutachtung eine Art „kritisches Zeitfenster“ existiert, innerhalb dessen wiederholte Befragungen durchgeführt werden sollten, um möglichst optimale Erinnerungsleistungen zu erzielen.

Zusammenfassend kann festgestellt werden, dass es in dieser Untersuchung kaum erkennbar einfacher oder schwieriger zu sein scheint, eine Aussage mit tatsächlichem Erlebnishintergrund als solche zu erkennen, als eine Falschaussage als solche zu identifizieren.

Dies entspricht nicht unbedingt den Ergebnissen aus Studien zur verhaltensorientierten Glaubwürdigkeitsattribution und widerspricht auch dem Befund einer generellen Tendenz, Aussagen eher als „wahr“ denn als „unwahr“ zu beurteilen (vgl. KÖHNKEN, 1990). Somit kann die Anwendung der kriterienorientierten Aussageanalyse in der forensischen Praxis als äußerst geeignetes Analyse- und Diskriminationsinstrument prinzipiell und auch offenkundig auch bei Anwendung auf spezifische Personengruppen wie die der Menschen mit Lernbehinderung nicht in Frage gestellt werden. Die CBCA ist als ein Teil einer umfassenden und die spezifischen Voraussetzungen der Person berücksichtigenden Methode zur Beurteilung der Glaubhaftigkeit einer Aussage (Statement Validity Analysis) zu sehen. Immer wieder müssen die Autoren der Kriteriologie selbst darauf hinweisen, dass die Ergebnisse einer merkmalsorientierten Inhaltsanalyse auf der Basis eines Aussagetranskripts ohne zusätzliche Informationen über die aussagende Person oder das relevante Ereignis für das abschließende Glaubhaftigkeitsurteil nicht ausreichen (vgl. STELLER, 1999). Auf die schon ausführlich dargestellte intraindividuelle Perspektive bei der Schlussfolgerung von der Qualität zur Glaubhaftigkeit einer Aussage soll hier nochmals hingewiesen werden. STELLER (1999) merkt an, dass zur Bewertung über die Trefferquoten gutacherlicher Urteile, Untersuchungen durchgeführt werden müssten, bei denen zusätzlich zu dem Interviewtranskript die gesamte relevante Information über die kognitiven, motivationalen, sozialen und anderen Fähigkeiten und Voraussetzungen zur Verfügung steht. Ein Aspekt der bei Simulationsstudien dieser Art zu bedenken ist, ist der der Komplexität der Inhalte der Aussagen. Die Jugendlichen konnten in Absprache mit der instruierenden Person zwischen mehreren verschiedenen Themen für ihre erlebnisbasierten und nicht-erlebnisbasierten Geschichten wählen. Für trainierte Rater ist es demnach generell schwieriger, intuitive Entscheidungsregeln zu entwickeln. Die Vielfalt der möglichen Themen ist auch für die Beobachterübereinstimmung von Bedeutung.

Demnach müssen in weiteren Forschungsarbeiten die eben genannten Aspekte bei der Auswahl der Versuchspersonen, deren individuellen intellektuellen Kompetenzen bzw. Defiziten im Sinne einer möglichst homogenen Stichprobe sowie einer bedachten Auswahl und ökologischen Validität von gegebenenfalls verwendetem Stimulusmaterial sorgfältig beachtet und umgesetzt werden, um sich trotz der methodisch-kritischen Umsetzbarkeit und Übertragbarkeit gewisser Aspekte und Teilbereiche von Simulationsstudien der forensischen Realität möglichst weit anzunähern.

5.3 Zusammenfassende und sonstige Aspekte

Da in der vorliegenden Untersuchung zur Gedächtnisleistung und Quellenidentifikation keine eigentliche Kontrollgruppe normal begabter Jugendlicher Teil des Versuchsdesigns war und die untersuchten Teilfragestellungen sehr umfassend und differenziert sind, ist - wie an anderer Stelle bereits erwähnt - eine direkte Bezugnahme auf alle Befunde bzw. Befunde von anderen Untersuchungen nicht immer möglich. Zum einen liegen diesbezüglich nur sehr wenige (Vergleichs-)Studien bei lernbehinderten Jugendlichen oder generell intelligenzgeminderten Personen vor, zum anderen wurden dort fast immer Vergleiche mit normal begabten Personengruppen angestellt. Im Übrigen bestand in einigen Studien das Stimulusmaterial aus Wortlisten oder anderen ökologisch nicht besonders validen Versuchsbedingungen. Daher sollen im Folgenden ergänzend zu den bisherig diskutierten Befunden allgemeine Erklärungsmodelle und eventuell relevante Ursachenaspekte angeführt werden.

Wenn möglich werden die aus der vorliegenden Arbeit gewonnenen Befunde so konkret wie möglich in die vorhandenen entsprechenden empirischen Befunde in Beziehung gesetzt.

Grundsätzlich kann nach MILNE & BULL (2001) bei lernbehinderten Menschen von einer generell verlangsamten Informationsverarbeitung ausgegangen werden, was sich unter anderem, wie auch in dieser Untersuchung in knappen und zunächst weniger vollständigen freien Berichten widerspiegelt. Im Vergleich zu gesunden Stichproben sind diese Schilderungen im Verhältnis jedoch nicht unbedingt mit mehr fehlerhaften Details versehen. In bisherigen Untersuchungen ließ sich bei lernbehinderten Versuchspersonen eine verstärkte Tendenz zu Intrusionen oder Konfabulationen in dem Sinn, dass zuvor nicht präsentierte Information Eingang in die Berichte der Versuchspersonen findet, nicht nachweisen. Insgesamt fanden sich keine signifikanten Korrelationen zwischen intellektueller Fähigkeit, verbalem Erinnerungsgedächtnis, Suggestibilität oder Zustimmungstendenz (GUDJONSSON & CLARE, 1995). Entscheidend hierbei ist offensichtlich, ob die Personen über einzelne Vorfälle oder verschiedene, mehrere Vorfälle eines wiederholten Ereignisses berichten (POWELL & THOMSON, 2003). Das größte Problem liegt nach Auffassung von MILNE & BULL (2001) eher im Abruf und der Wiedergabe der gespeicherten Informationen, was durch Schwierigkeiten in Sprachverständnis und –ausdruck noch verstärkt wird.

Dies konnte bezüglich des Vorkommens von Intrusionen und Konfabulationen in der vorliegenden Untersuchung nicht so pauschal bestätigt werden. Wie ausführlich dargestellt, ergaben sich zu den einzelnen Befragungszeitpunkten hinsichtlich der Dia- bzw. Unfall-Serie 'PKW-Fahrrad' durchaus jeweils große Unterschiede zwischen den Versuchsgruppen.

Auch Benetto et al. (1996, zit. nach LORSBACH, 2000) fanden bei autistischen Jugendlichen bei Erinnerungsaufgaben mehr Fehler in Form von Intrusionen als bei nicht-autistischen, wonach die autistischen Jugendlichen ein Defizit hinsichtlich der Quellenzuordnung hätten und nicht oder nur schwer in der Lage sind, den Kontext der aktuellen Aufgabe zu nutzen, um ungeeignete Antwortalternativen auszuschließen.

Im Zusammenhang mit der Diskussion des Vorkommens schematypischer Intrusionen ist es nach GREENBERG et al. (1998) dann wahrscheinlicher, sich eher auf Skripte und Schemata als auf die ursprüngliche Gedächtnisspur zu verlassen, wenn Elemente aus dem Skript zentral bzw. mit dem Skript selbst eng verbunden sind und je länger das Behaltensintervall ist. Dabei können Schemata wie Skripte sowohl positive wie negative (z. B. falsche Schlussfolgerungsprozesse) Effekte haben, Interpretationen beeinflussen und Elaborationen ermöglichen. Im späteren Verlauf, wenn Erinnerungen bereits unklar und verschwommen sind, haben Menschen somit die Möglichkeit sich auf das Skript zu beziehen um Gedächtnis- und Erinnerungslücken zu schließen. Dies stellt auf der anderen Seite jedoch auch eine Fehlerquelle vor allem in Form von skriptbasierten Intrusionen dar. Man geht heute davon aus, dass schemadiskrepante Details kurzfristig besser und langfristig schlechter erinnert werden als schemakonsistente Details. Diese Annahmen konnten bei Betrachtung der Ergebnisse der Vergleiche T1-NT2 bzw. T2-NT2 der vorliegenden Arbeit nur teilweise bestätigt werden. Nach GREUEL (2001) fällt zudem bereits die Enkodierung und genaue Rekonstruktion distinkter Episoden umso schwerer, je unvertrauter und komplexer das übergeordnete Ereignisschema ist. Insgesamt tragen nach KELLOWAY et al. (2004) Einstellungen, Skripte, Schemata und Stereotype möglicherweise zu Erinnerungsfehlern bei.

In der Studie von CLARE & GUDJONSSON (1993) ergab sich insgesamt betrachtet, dass die Versuchspersonen mit einer Intelligenzminderung in allen untersuchten Bereichen (Suggestibilität, Konfabulationen, Zustimmungstendenz) sich diesbezüglich als anfälliger erwiesen als die normalbegabten Personen aus der Kontrollgruppe und zudem schlechtere Erinnerungsleistungen zeigten. Zudem fanden sich Konfabulationen und Zustimmungstendenzen bei Menschen mit einer Lernbehinderung signifikant häufiger als bei Menschen ohne eine intellektuelle Beeinträchtigung. Die Tendenz, Fragen bejahend zu beantworten, korrelierte negativ mit der allgemeinen intellektuellen Fähigkeit. Zudem stellte sich heraus, dass schlechtere Erinnerungsleistungen mit einer höheren Anzahl an Konfabulationen korrelierten. Die Autoren schlussfolgern daraus, wenn Menschen mit Lernbehinderung in höherem Ausmaß konfabulieren als andere mit durchschnittlichen Fähigkeiten, dann machen sie selbst unter optimalen Befragungsbedingungen weniger zuverlässige Aussagen.

Bezüglich der für Zeugenaussagen zentralen Fähigkeit zwischen mehreren unterschiedlichen Informationsquellen unterscheiden bzw. diese voneinander abgrenzen zu können (internalen wie externalen) sei neben den zwei bisher diesbezüglich zwei bekannten Studien bei lernbehinderten Kindern (u. a. LORSBACH & EWING, 1995) auf die Befunde der vorliegenden Arbeit verwiesen. Zunächst stellen LORSBAH & EWING (1995) wiederum im Vergleich mit normalbegabten Kindern dazu fest, dass lernbehinderte Kinder, die an Quellenidentifikationsexperimenten teilnahmen, sowohl im Wiedererkennensteil, aber auch im Quellenidentifikationsteil deutlich schlechter abschneiden als die Kontrollgruppe. Dies mag zum Großteil daran liegen, dass die oft komplexen Kontextinformationen entweder nicht spontan enkodiert wurden und/oder nicht systematisch abgerufen werden können (vgl. LORSBACH, 2000). In beiden Bedingungen (‘internal source monitoring’ und ‘reality monitoring’) sind verschiedene Gedächtnisleistungen und Entscheidungsprozesse gefordert, respektive der Einordnung und Verarbeitung sowohl sensorischer Attribute als auch damit verbundener kognitiver Operationen. Nach LORSBACH (2000) können Gedächtnisprobleme lernbehinderter Kinder zwar nicht verallgemeinert werden, jedoch sind diese offenbar auf Aufgaben beschränkt sind, die sich auf bewusste Erinnerungsversuche beziehen. Ähnlich wie bei jüngeren Kindern haben diese Schwierigkeiten bei kognitiv anspruchsvollen sowie prozesskontrollierten Aufgaben, allerdings nicht bei automatisch auszuführenden Aufgaben oder Entscheidungsprozessen. Hierbei zeigten sich nach neuesten Erkenntnissen nur im Fall expliziter Gedächtnistestung Unterschiede zu normalbegabten Kindern, nicht wenn die Testung implizit verläuft (vgl. LORSBACH, 2000).

In der vorliegenden Arbeit mit seinem anderen, auch sehr komplexen und differenzierten Design konnten zwar einerseits unabhängig von allen drei Befragungszeitpunkten keinerlei wesentlichen Unterschiede zwischen den beiden Versuchsgruppen festgestellt, was mögliche Quellenverwechslungen insgesamt wie auch innerhalb der Serien betrifft; andererseits finden sich zu allen drei Befragungszeitpunkten bei Abfrage der Dia-Serie ‘PKW-PKW’ in den Berichten derjenigen Jugendlichen, die zwischen zwei externalen Quellen unterscheiden mussten, deutlich mehr Quellenverwechslungen als in den Schilderungen der Jugendlichen, die zwischen einer externalen und einer internalen Quelle differenzieren sollten.

Die Genauigkeit von Quellenzuordnungen bzw. die Quellenidentifikation ist nach JOHNSON & RAYE (2000) generell in Simulationsstudien wie der vorliegenden von mehreren Faktoren abhängig, wie etwa Charakteristika der Frageformulierung oder Interviewtechnik, verschiedenen Graden der verfügbaren und vorhandenen Information der aktivierten Erinnerungspfade, der Einschätzung der Menge an perzeptuellen Details, der Vertrautheit und Spezifität und generell den jeweiligen Anforderungen an die entsprechende Situation. Daher basiert

die Qualität der Quellenidentifikation fundamental auf den Informationen, die anfangs über das Ereignis aufgezeichnet wurden (JOHNSON & MULTHAUP, 1992). Somit ist ganz besonders die Qualität der Enkodierung eines Ereignisses oder einer Beobachtung von entscheidender Bedeutung. Alle Vorgänge - die das Ergebnis von spezifischen visuellen und kognitiven Prozessen sind, die während einer Wahrnehmung wirksam sind - und die die Person davon abhalten, die gesamten Kontextinformationen aufzunehmen, reduzieren somit die Enkodierung von potentiell relevanten Quelleninformationen. Bekannte Faktoren wie das Erregungsniveau, selektive Aufmerksamkeitsfokussierung oder bestimmte soziale Erwartungen beeinflussen ebenso wie die angesprochenen Quellen(-kombinationen) als Distraktoren unmittelbar die Entscheidungsprozesse (vgl. GREUEL, 2001). Allgemein lässt sich festhalten, dass geteilte oder selektive Aufmerksamkeit in bestimmten Situationen dazu führt, dass sich Personen nicht mit zusätzlichen Prozessen beschäftigen können, die nötig wären, um eine Information fest zu etablieren und für die Identifizierung der korrekten Quelle heranzuziehen. So wird nach Lindsay wie auch Hekkanen & McEvoy in erster Linie die perzeptuelle Ähnlichkeit zwischen den Erinnerungsquellen als einer der Hauptfaktoren für Quellenzuschreibungsfehler diskutiert.

Fehler treten nach der Quellenidentifikationshypothese dann auf, wenn die irreführten Versuchspersonen fälschlicherweise die Erinnerung an die Falschinformation als Erinnerung an das beobachtete Ereignis identifizieren. Da eine interrogative Suggestibilität bzw. ein Einfluss einer Falschinformation während der Befragung durch den Befrager eigentlich weitestgehend ausgeschlossen werden kann, bleibt die Frage offen, inwieweit, ob und wann welche Art eines 'post-event-misinformation-effects' sich auf die Aussagequalität der Versuchspersonen ausgewirkt hat. Nach Meinung des Verfassers sind allerdings intensive Absprachen oder der Austausch der Jugendlichen untereinander über das Stimulusmaterial selbst eher als gering einzustufen, können natürlich jedoch nicht ausgeschlossen werden.

Weiterhin weist LORSBACH (2000) darauf hin, dass bei Kindern mit intellektuellen Defiziten vor allem das explizite Gedächtnis gestört ist, was jedoch den Gebrauch von reflektiven Prozessen erfordert und entscheidend für den Aufbau differenzierter perzeptueller wie semantischer Informationen und kognitiver Operationen ist, welcher sich wiederum insbesondere bei Quellenidentifikationsaufgaben als kritisch und problematisch erweist.

Die Befunde der vorliegenden Arbeit decken sich im Rahmen der Vergleichbarkeit der Untersuchungen mit den Ergebnissen von Lorschbach et al. (1992) sowie mit der Überzeugung und der allgemeinen Ansicht, dass auf bildhaften Repräsentationen basierende Erinnerungen zugänglicher und leichter abzurufen sind als solche, die auf auditorischen oder verbalen Repräsentationen beruhen (*picture superiority effect*). Insgesamt hängt eine problemlose und detaillierte Quellenzuordnung von der Menge und Qualität der ursprünglich wahrgenomme-

nen Information ab, wobei sich Qualität diesbezüglich auf den Grad der Unterscheidbarkeit der Informationen bezieht. Dabei sind sowohl perzeptuelle als auch kontextuelle Einzelheiten von Bedeutung. Die bei lernbehinderten Personen gut dokumentierten Enkodierungsprobleme, also der mangelnden Fähigkeit, komplexere und in hohem Maße differenzierbare Gedächtnisrepräsentationen zu bilden, resultieren somit möglicherweise in einer der Person zur Verfügung stehenden verminderten Menge und Spezifizierung der Information über die Quelle der unterschiedlichen Gedächtnisinhalte. Werden jedoch verschiedene Ereignisse präsentiert, die charakteristische Merkmale aufweisen (z.B. Bilder gegenüber gesprochenen Worten), sind die Gedächtnisinhalte auch für lernbehinderte Kinder leichter zu unterscheiden, was sich auch in zuverlässigeren und genaueren Quellenzuordnungen manifestiert (vgl. LORSBACH, 2000).

Offensichtlich ist nach SUSSMAN (2001) ein gradueller Entwicklungsanstieg hinsichtlich der Realitätsüberwachung von vorgestellten Ereignissen vom Vorschulalter bis in die Jugendzeit zu verzeichnen. Demnach scheint die Fähigkeit sich mit Realitätsüberwachung auseinanderzusetzen mit dem Alter von 11 oder 12 Jahren vollständig entwickelt. Inwieweit diese Befunde für lernbehinderte Jugendliche oder generell Menschen mit Intelligenzminderung gelten oder übertragbar sind, bleibt auch aufgrund der Individualität und Komplexität der Störung in dieser Untersuchung zumindest kritisch nicht gänzlich klar zu beantworten, da unter anderem keine Informationen vorlagen, ob und welche Versuchspersonen der ca. 18-jährigen Jugendlichen ein entsprechendes Entwicklungsalter erreicht haben. Insgesamt kann dies jedoch aufgrund des Eindrucks aus den vielen Befragungssituationen tendenziell doch angenommen werden.

Obwohl die Zahl der Studien zur Quellenidentifikation begrenzt ist, kann nach dem gegenwärtigen Forschungsstand davon ausgegangen werden, dass lernbehinderte Kinder Schwierigkeiten haben, bewusst Gedächtnisquellen zu erinnern, die verbale Informationen beinhalten. Daraus kann jedoch nicht einfach auf ein grundsätzliches Gedächtnisdefizit geschlossen werden. Vielmehr deuten die Befunde eher auf ein grundlegendes Problem im Gebrauch der von JOHNSON et al. eingeführten Begriffe und Phänomene der heuristischen und systematischen Entscheidungsprozesse hin (vgl. LORSBACH, 2000). In diesem Zusammenhang ist es allerdings wichtig darauf hinzuweisen, dass es sich bei einigen der genannten Studien zum einen um experimentelle Paradigmen handelt, zum andern nicht besonders viele Untersuchungen dazu existieren. Allgemeingültige und schlussfolgernde Befunde gestalten sich demnach zum gegenwärtigen Zeitpunkt der Forschung noch schwierig. Inwiefern bzw. ob sich Leistungsniveaus in forensisch relevanten Situationen aufgrund persönlicher Involviertheit, emotionalem Bezug oder anderen sozialen und motivationalen Faktoren verbessern

oder verschlechtern, kann und muss auch als Funktion der Kompetenz des Befragenden verstanden werden.

Unter paralleler Beachtung der Befunde von beispielsweise PERLMAN et al. (1994), GUDJONSSON & CLARE (1993) oder auch MILNE & BULL (2001) bezüglich begangener Fehler (falsch oder nicht mehr erinnelter Details) konnte in der vorliegenden Untersuchung hinsichtlich der insgesamt sowie im Bereich der allgemeinen Erinnerung (-sfehler) begangenen Fehler festgestellt werden, dass Jugendliche die zwischen einer externalen und einer internalen Quelle differenzieren müssen, nach acht Wochen in beiden Serien jeweils (d.h. bei wiederholter und auch erster Befragung) wesentlich mehr Fehler begehen als diejenigen Jugendlichen, die zwischen zwei externen Quellen unterscheiden müssen. Dieser Befund gilt teilweise nur tendenziell auch für einen ersten Befragungszeitpunkt T1. Somit ist es für Versuchspersonen, die zwischen zwei semantisch ähnlichen externalen Quellen unterscheiden müssen, offensichtlich einfacher, weniger fehlerbehaftete Schilderungen dieser Ereignisse abzugeben.

Allgemein verweisen McNAMARA & WONG (2003) unter Bezugnahme auf einerseits die Hypothese der Zugänglichkeit von Informationen und einer bestimmten Wissensform, dem 'general event knowledge' und Swanson (1984) darauf, dass lernbehinderte Schüler möglicherweise Defizite in der ausführenden Komponente des Arbeitsgedächtnisses haben, die allerdings unabhängig von sprachspezifischen Verarbeitungsproblemen sind. Schlechte Erinnerungsleistungen könnten demnach entweder auf das Fehlen einer passenden Gedächtnisspur (*storage deficiency*) oder Schwierigkeiten beim Zugang während des Behaltensvorgangs (*retrieval deficiency*) zurückzuführen sein. Eine derartige Unterscheidung ist jedoch problematisch so zu treffen, da es methodisch schwierig ist, Speicherungs- bzw. Enkodierungsprozesse von Behaltensprozessen zu trennen. Swanson (1984; zit. nach McNAMARA & WONG, 2003) fand, dass lernbehinderte Schüler beim Erinnern nonverbaler, komplexer visueller Informationen bessere Leistungen erzielten als eine Kontrollgruppe. Die Befunde deuten auf die ausgeführten Erklärungsmodelle hinsichtlich von fehlerhaften Prozessen des Arbeitsgedächtnisses sowie dem ungenügenden und nicht effektiven Gebrauch von Behaltensstrategien ebenso wie eines Abrufdefizits bezüglich bereits früher enkodierter Information hin.

Dabei kann davon ausgegangen werden, dass einige der von PRESSLEY (1991) genannten Basis-Kompetenzen einer intakten Informationsverarbeitung bei lernbehinderten Personen gestört bzw. zumindest wesentlich eingeschränkt sind. Insbesondere sei hier auf das defizitäre Vermögen des Gebrauchs von effektiven Strategien bei gegebenenfalls entsprechend vorhandenen detaillierten mentalen Repräsentationen sowie der mangelnden Fähigkeit zur angemessenen und selektiven Aufmerksamkeit und Konzentration bezüglich Informationen und deren Relevanz hingewiesen. Speziell im Bereich des Gedächtnisses sind hier Strate-

gien der Wiederholung, Organisation sowie elaborative Strategien unzureichend entwickelt wie auch im Bereich des Begriffsvermögens die Bildung repräsentativer Bilder und eine qualitativ und quantitativ ausgeprägte Aktivierung relevanten Vorwissens. Daneben können Angst hinsichtlich schulischer oder beruflicher Angelegenheiten sowie im Falle einer gegebenenfalls vorhandenen Wissensbasis Probleme des Zugangs und der Anwendung auftreten.

Die Befunde von WINOGRAD (1998) wie auch die der vorliegenden Arbeit lassen die Frage offen, in welchem Ausmaß individuelle Faktoren hinsichtlich einer bildhaften und mentalen Repräsentationsfähigkeit bei Studien zur Erinnerungsleistung und der Unterscheidung von Informationen und deren Herkunft bei Personen mit einer Lernbehinderung von Bedeutung sind und berücksichtigt werden müssen. Klar sein dürfte jedoch, dass der genannte Personenkreis durchaus Schwierigkeiten hat, intern mit Hilfe entsprechender Vorgaben generierte Quellen abzurufen, ohne dabei sehr hohe Fehlerraten zu produzieren.

Im aussagepsychologischen Kontext erscheint nach GREUEL (2001) die alleinige Anwendbarkeit der Heuristik „Lebhaftigkeit“ einer Aussage auf dem Hintergrund neuerer Studien problematisch, da hier wichtige Zusammenhänge zwischen individueller Imaginationsfähigkeit und Suggestibilität bestätigt werden konnten. Demnach verringert sich bei Personen mit hoher Imaginationsfähigkeit nicht nur die phänomenale Diskrepanz zwischen Erlebniserinnerung und Vorstellung, so dass die Gefahr von Quellenzuschreibungsfehlern erhöht ist. So kommt es auch zu einer erhöhten Anfälligkeit für Suggestionen, so dass reine Vorstellungsinhalte unter spezifischen Bedingungen auch fälschlicherweise als Kindheitserinnerungen anerkannt und übernommen werden können. Insofern erweist sich die intuitiv nahe liegende Strategie, allein anhand der Lebhaftigkeit der Erinnerung auf deren Wahrheitsstatus zu schließen, als unzuverlässig. Die Befunde der vorliegenden Arbeit zur Quellenüberwachung bei lernbehinderten Jugendlichen geben einen ersten Hinweis, wie sich die Zusammenhänge zwischen imaginärer Vorstellungskraft und der Wahrscheinlichkeit von Quellenzuschreibungsfehlern bei dieser speziellen Personengruppe darstellen. Jedoch müssen diese Befunde in einigen weiteren Forschungsarbeiten empirisch weiter intensiv geprüft und validiert werden. In jedem Fall scheinen hier andere Muster und Prozesse als bei Kindern bestimmter Altersgruppen oder normalbegabten Erwachsenen zu gelten. Bezüglich normalbegabter Versuchspersonen und im Sinne eines ‘Stimmungs-Kongruenz-Effekts’ ergaben sich erste Anhaltspunkte dafür, dass die Qualität der affektiven Informationen Einfluss auf die Quellenzuordnung hat.

Bisherige Ergebnisse aus entsprechenden Studien deuten darauf hin, dass Kinder mit einer Intelligenzminderung Informationen genauso gut enkodieren, speichern und abrufen können wie Kinder des gleichen Entwicklungsalters, sofern es sich nicht um künstlich erzeugte Gedächtnisexperimente handelt. Ein automatisiertes Erinnern scheint leichter für diese

Personengruppe zu sein, da es keine Gedächtnisstrategien wie z.B. inneres Vorsagen benötigt (vgl. HENRY & GUDJONSSON, 1999). Im Rahmen der genannten wie auch der vorliegenden Untersuchung ist jedoch kritisch zu beachten, dass die Gruppe der intellektuell beeinträchtigten Kinder bzw. Jugendlichen sehr unterschiedliche Intelligenzniveaus aufwies; so weist etwa die Hälfte der Kinder eine geistige Behinderung auf, die restlichen Kinder sind lernbehindert. In der vorliegenden Untersuchung ist die Spannweite des Intelligenzniveaus ebenfalls recht groß, was bei einigen der Jugendlichen streng genommen - bei reiner Orientierung am kognitiven Niveau - auch nicht mehr den ohnehin kritischen Begriff der 'Lernbehinderung' umfassen würde.

Zwar verweisen MILNE (1999) und BULL (1995) darauf, dass es einer Befragung im juristischen Kontext durchaus sinnvoll und fördernd sein könnte, Zeugen über das Vorgehen bei Gericht genau zu informieren, ihnen den Zweck der Befragungen zu erklären und herauszustellen, dass die Interviewenden selbst nichts über die von intellektuell beeinträchtigten Zeugen zu bekundenden Vorfälle wissen, da diese bei dem fraglichen Ereignis nicht anwesend waren und somit nicht wüssten, was passiert sei. Andernfalls besteht zumindest bei Kindern mit einer Lernbehinderung die Möglichkeit, dass diese die Vermutung haben, der Interviewer sei kundig. Jedoch ist dies auf dem wissenschaftlichen Terrain auf dem man sich mit Simulationsstudien gewöhnlich bewegt und die - auch wenn sie hinsichtlich Aspekten der ökologischen Validität möglichst optimiert sind (wie hoffentlich auch die vorliegende Arbeit) - eben doch nicht gänzlich auf die Realität übertragbar sind. So sollte dennoch in künftigen Forschungsarbeiten nichts unversucht bzw. unbedacht bleiben, um sich zumindest konstant einer möglichst realitätsnahen Simulation der forensischen Realität zu nähern, um möglichst zuverlässige Anhaltspunkte für den Umgang und der Auseinandersetzung vor allem mit spezifischen Personengruppen zu gewinnen.

Mit Bezugnahme auf die Ergebnisse von D'ARGEMBEAU et al. (2003) kann allgemein festgestellt werden, dass emotionale Geschehnisse besser, d.h. facettenreicher enkodiert werden als neutrale, da dass positive (autobiographische wie wohl auch episodische) Erinnerungen mehr sensorische und kontextuelle Einzelheiten aufwiesen als negative oder neutrale, wobei sich negative und neutrale Erinnerungen in den meisten Dimensionen nicht besonders unterschieden. Andererseits ist hierfür auch ein höheres Maß an Aufmerksamkeit erforderlich. Jedoch werden diese Prozesse und Informationen diesbezüglich quasi freiwillig eher bewusst und sorgfältiger elaboriert, da emotional bedeutsame Ereignisse von größerem persönlichen Interesse sind als neutrale. Zudem wird an emotional Bewegendes öfter gedacht und dies auch meistens öfter anderen Personen mitgeteilt, was sich demnach auch in sich positiv auswirkenden, spezifischen neuronalen Konsolidierungsprozessen äußert. Außerdem fanden die Autoren, dass emotionale und neutrale Ereignisse aus unterschiedlichen Perspektiven heraus erinnert werden und Personen, die Gefühle eher unterdrücken

sich mit Ausnahme von qualitativen Merkmalen ihrer negativen Erinnerungen diesbezüglich nicht von anderen Menschen unterscheiden. So ist hinsichtlich der vorliegenden Teiluntersuchung zur Gedächtniskompetenz und Quellenidentifikation trotz einer relativ hohen ökologischen Validität darauf hinzuweisen, dass wohl für die meisten der Versuchspersonen die präsentierten Dia-Serien eines offenkundig gestellten Verkehrsunfalls ohne große Verletzungen eher im emotional weniger aufrüttelnden, um nicht zu sagen sehr neutralen Bereich anzusiedeln sind. Zwar stellte für einige der Jugendlichen die gesamte Stimuluspräsentation, die Instruktionen und Befragungstermine eine nette, neue Abwechslung in ihrem Alltag dar, jedoch fehlte wohl meistens der direkte persönliche und intensive Bezug. So ist es durchaus vorstellbar und nachzuvollziehen und war auch bereits während der Dia-Präsentation bei manchen Jugendlichen zu beobachten, dass wenig Interesse und in der direkten Konsequenz auch weniger Aufmerksamkeit vorhanden war, was sich wiederum in schlechteren Enkodierungsleistungen widerspiegelt. Aus methodisch schwierigen Gründen kann generell und auch an dieser Stelle nicht vollständig geklärt werden, welcher Anteil an fehlerhaften oder ungenauen Zeugenaussagen jeweils unzureichenden Enkodierungs- oder Dekodierungsprozessen zukommt oder in welchem Ausmaß die Speicherung bzw. der Abruf von Informationen gestört ist. Zum einen ist dies methodisch eben nicht gut zu trennen, zum anderen ergeben sich gerade hier vermutlich so große individuelle Unterschiede, dass eine konkretes Befundmuster oder ein genereller Zusammenhang wohl schwer zu bestimmen sein wird.

Eine weitere Überlegung besteht nach Meinung des Verfassers der Arbeit darin, ob es in Analogie zum beschriebenen Phänomen des 'weapon focus' auch in „harmlosen“ bzw. relativ neutralen Simulationsstudien ein vergleichbares Phänomen bezogen auf neutralere Objekte oder Personen (eine Art 'car focus' bzw. 'Personen-Fokus') gibt bzw. inwiefern sich gegebenenfalls auch geschlechtsspezifische Unterschiede auswirken. So wäre es denkbar, dass möglicherweise einige Jugendliche (eventuell mehr auf Seiten der männlichen Versuchspersonen) im Rahmen der Enkodierung eher bzw. in einem so großen Ausmaß auf die Marken der PKW's oder des Fahrrades konzentrierten, dass eine zusätzliche Fokussierung auf beteiligte Personen oder deren Haarfarbe oder Bekleidung die kognitive und/oder motivationale Kompetenz dieser überschritten hat, insbesondere aufgrund der ohnehin eingeschränkten Kapazitäten vor allem im Bereich der Informationsverarbeitung von peripheren, perzeptuellen Details. Umgekehrt wäre es denkbar, dass etwa einige der weiblichen Versuchspersonen (ohne an dieser Stelle Vorurteile oder andere assoziative Feuer zu schüren zu wollen) eher auf bestimmte Aspekte der Kleidung oder des Kinderwagens als auf Farben oder Fabrikate von Autos oder anderen für sie vermeintlich nicht so relevanten Details achteten und sich somit jeweils in bestimmten Bereichen entsprechender Zeugenaussagen gewisse Schwerpunkte bzw. Fehler finden.

Kinder, Erwachsene wie wohl auch lernbehinderte Jugendliche erinnern persönlich bedeutsame Ereignisse am besten. Dabei gilt das kindliche Gedächtnis als idiosynkratisch, so dass für Kinder gänzlich andere Aspekte bzw. Ereignisse bedeutungsvoll und damit erinnerungswürdig sind als für Erwachsene. Die Behaltensleistung an sich hängt bei Kindern wie Erwachsenen im Wesentlichen von der Distinktheit des fraglichen Ereignisses ab, wobei hier in erster Linie das Ausmaß der persönlichen Bedeutsamkeit entscheidend ist. Inwieweit die genannten Annahmen auf Personen mit Intelligenzminderung zutrifft bzw. zu welchem Grad die Gedächtnisleistung auch von vom spezifischen Sachverhalt, besonderen Interessen, motivationalen Komponenten oder dem in dieser Untersuchung verwendeten Versuchsmaterial abhängt, konnte und kann an dieser Stelle nicht endgültig geklärt werden und verlangt nach weiteren Untersuchungen.

Hinsichtlich des Vergleichs von über die Zeit konstanten, generellen Aussageinhalten kann bezüglich beider Teiluntersuchungen der vorliegenden Arbeit Folgendes festgehalten werden: Für die Bereiche des zentralen Kerngeschehens, eigener Rollen (Aktivitäten) und beteiligter Handlungspartner sind eher konstante Aussageinhalte und weniger Fehler zu erwarten. Diese müssen auch für intentional falsch aussagende Zeugen weitestgehend konstant sein, da diese eher schemarelevant sind. Anders dagegen verhält es sich hinsichtlich der Aspekte von Örtlichkeiten, unmittelbar handlungsrelevanten Gegenständen, Lichtverhältnissen und globalen Körperpositionen beteiligter Personen. Diese können jeweils variieren (vgl. Konstanzmodell nach ARNTZEN, 1993).

Im Rahmen selbst erlebter oder als Augenzeuge direkt persönlich beobachteter Geschehnisse können auch so genannte periphere Aspekte zentral sein, für „Lügner“ dagegen eher nicht.

Insgesamt kann auch angenommen werden, dass autobiographische Erinnerungen an persönliche, emotional bedeutsame Erlebnisse im Allgemeinen zuverlässig sind, insbesondere was deren zentralen, aktionalen Gehalt betrifft. So treten unter adäquaten Abrufbedingungen (freier Bericht) in erster Linie Auslassungsfehler auf, so dass eher von unvollständigen als von unzuverlässigen Erinnerungsberichten auszugehen ist.

Zusammenfassend kann aufgrund einiger angeführter Studien zur Zeugenkompetenz bei Menschen mit einer Intelligenzminderung sowie einigen Befunden der vorliegenden Arbeit angenommen werden, dass von einer generellen Zeugenuntüchtigkeit nicht ausgegangen werden kann. Defizite sind nur dann offensichtlich geworden, wenn gleichaltrige Personen den Maßstab bilden. In den zitierten Studien, bei denen Menschen im selben Entwicklungsalter die Vergleichstichprobe darstellen, konnten keine manifesten Unterschiede gefunden werden.

5.4 Ökologische Validität sowie Implikationen und Ausblick für zukünftige Forschungsarbeiten und die forensische Praxis

Inwieweit die Ergebnisse der vorliegenden Arbeit auf die forensische und psychologische Praxis anwendbar sind, hängt in sehr hohem Maß von der ökologischen Validität des Untersuchungsdesigns ab (Orientierung des experimentellen Vorgehens an der Realität). Dies stellt seit jeher ein grundsätzliches Problem der experimentellen Forschung vor allem im Bereich der Aussage- und forensischen Psychologie dar, besonderes auch im Hinblick auf die Qualität der Aussagen sexuellen Missbrauchs, da einerseits möglichst gut kontrollierte, andererseits aber auch möglichst ökologisch valide Untersuchungen durchgeführt werden sollen.

Ein Feldexperiment weist naturgemäß immer eine höhere ökologische Validität als eine - wie auch die vorliegende Untersuchung – Simulationsstudie mit der Darbietung von Diaserien, da verschiedene Faktoren auf Zeugen einwirken, denen im experimentellen Setting Grenzen gesetzt sind und die nicht nachvollzogen werden können. Jedoch wurde beispielsweise bei der Durchführung der Interviews mit den Versuchspersonen darauf geachtet, sie Befragungen von Polizei, Richter oder Gutachter anzugleichen (zumindest wie sie theoretisch ablaufen sollten, es aber leider in der Realität oft nicht sind), indem zunächst ein freier Bericht des Geschehens abgefragt wurde, gefolgt von gezielten Nachfragen (Trichtertechnik). Des weiteren wurde darauf geachtet, eine Analogie zur Praxis herzustellen, indem zwischen den beiden Befragungszeitpunkten ein Delay von ca. acht Wochen lag; da in der forensischen Praxis ebenfalls häufig große zeitliche Abstände zwischen der Beobachtung eines Ereignisses, der ersten Befragung und den nachfolgenden Vernehmungen liegen, wurden die Interviews in der vorliegenden Arbeit zu zwei Befragungszeitpunkten in einem zeitlichen Abstand von sieben bzw. acht Wochen durchgeführt. Es ist jedoch anzumerken, dass die Befragung von Versuchspersonen in experimentellen Untersuchungen wie der vorliegenden vermutlich einfacher ist als tatsächliche Vernehmungen von Zeugen. Die Versuchspersonen dieser Studie nahmen zum einen an dem Experiment freiwillig teil, andererseits wurden sie zu einem relativ neutralen Thema, nämlich zu Verkehrsunfällen, befragt. Die emotionale Belastung oder sogar die Traumatisierung ist bei Personen, die tatsächlich Zeuge und/oder Opfer eines schwerwiegenden Ereignisses wurden, meist erheblich höher. Dies führt dazu, dass sich reale Befragungssituationen meist erheblich schwieriger gestalten als die Interviews in der hier durchgeführten Untersuchung, da Zeugen häufig voreingenommen oder skeptisch gegenüber denjenigen Personen sind, die sie vernehmen. So gilt es generell in zukünftigen Untersuchungen einerseits die Grenzen der ökologischen Validität bzw. einer möglichst realitätsnahen Simulation maximal auszureizen bzw. sich dieser anzunähern. Andererseits dürfen die Grenzen der ethischen Zumutbarkeit natürlich ebenfalls nicht überschritten wer-

den. So muss auch in diesem Zusammenhang ein für alle Beteiligten und dem Fortschritt wissenschaftlicher Erkenntnisse angemessener Kompromiss gefunden werden, in dem die genannten Aspekte vernünftig gegeneinander abgewogen sind.

Die bisherigen Befunde entsprechender Studien, lassen den Schluss zu, dass sich intellektuell beeinträchtigte Kinder beeinflussbarer zeigten durch die antizipierte Unangemessenheit ihrer ursprünglichen Antworten bedingt durch eine erneute, gleichartige Befragungssituation. Die vermehrten Änderungen im Aussageverhalten ('eyewitness shift') stellen etwa nach Ansicht von Henry & Gudjonsson scheinbar eine besondere Schwäche von Kindern mit einer Intelligenzminderung da. Demnach sind Mehrfachbefragungen bei Zeugen zumindest kritisch bzw. problematisch anzusehen. Gudjonsson und Clark, führen zwei Gründe an, warum Menschen mit einer Intelligenzminderung in einem erhöhten Maße suggestibel sind: Zum einen durch eine schwächere Gedächtniskapazität ('kognitiver Faktor') und zum anderen durch eine geringere Fähigkeit, sich vor Unsicherheiten und Erwartungen bei Befragungen zu schützen ('sozialer Faktor') (HENRY & GUDJONSSON, 2003). Interessanterweise konnten die negativen Konsequenzen einer wiederholten Befragung unabhängig von einer intellektuellen Beeinträchtigung beobachtet werden. Insgesamt bleibt jedoch noch offen, inwieweit die Ergebnisse auch auf reale, strafrechtlich relevante, beobachtete Handlungen übertragbar sind. Daher sollten für künftige Forschungsarbeiten aus der vorliegenden Arbeit einige Implikationen und Anregungen ableitbar sein.

So sollten einerseits in nachfolgenden Untersuchungen homogenere, möglichst präzise definierte Stichproben bezüglich intellektueller Kompetenz und kognitiver Einschränkung ausgewählt bzw. zusammengesetzt werden, um die interne Vergleichbarkeit der Studie und deren Ergebnissen zu optimieren. Andererseits sollten studienübergreifend auch heterogenere Stichproben als die eigentlich schon sehr spezifische Personengruppe der jugendlichen Lernbehinderten untersucht werden, da Zeugen von strafrechtlich relevanten Geschehnissen bzw. auch Opfer von Straftaten Personen aller Bevölkerungsschichten, jedes Alters oder sonstiger Charakteristika sein können. Je heterogener die Stichprobe, desto eher entspricht sie auch der Grundgesamtheit. Somit besteht diesbezüglich noch ein großes Spektrum an möglichen zu untersuchenden Personengruppen und interessanten Fragestellungen sowie im Sinne einer Optimierung des Operschutzes und des Umgangs mit kritischen Zeugengruppen in der forensischen Realität.

Weiterhin ist es nach Müller-Luckmann wichtig zu erwähnen, dass die Kompetenzfrage durchaus situationsspezifisch zu beurteilen ist. So können minderbegabte Zeugen beispielsweise Schwierigkeiten haben, einen als Augenzeuge erlebten, komplexen Verkehrsunfall zu

schildern, andererseits aber durchaus in der Lage sein, autobiographische Erinnerungen korrekt abzurufen und zu schildern oder glaubhafte Aussagen über ein selbst erlebtes Delikt gegen die sexuelle Selbstbestimmung zu produzieren.

Im Allgemeinen nahm in der Studie von GORDON et al. (1994) die Erinnerungsfähigkeit der befragten Kinder nach sechs Wochen ab, wobei die Forscher im Hinblick auf zukünftige Forschung längere Befragungsverzögerungsintervalle vorschlagen, da dies eher der forensischen Praxis und Realität entspreche. Trotzdem sich in der vorliegenden Untersuchung aufgrund schon diskutierter Lerneffekte und der Gewöhnung an den Interviewer, die Testsituation und deren Ablauf sowie sich andeutender Reminiszenz- und Hypermnesie-Effekten die Ergebnisse hinsichtlich korrekt erinnerter Details nicht so ganz eindeutig bestätigen lassen, kann sich dennoch der oben genannten Tendenz des Vergessens über die Zeit hinweg angeschlossen werden wie auch dem Vorschlag - im Sinne einer möglichst realen Simulation - die zeitlichen Abstände im forensischen Prozedere noch länger zugestalten. Generell ist die Diskussion zum Thema Glaubhaftigkeit und Erinnerungsverlust immer kritisch zu sehen, insofern, dass wenn ein Detail erst zu einem zweiten oder dritten Befragungszeitpunkt auftaucht, spricht dies nicht unbedingt gegen die Glaubhaftigkeit der Erinnerung. So kann und muss dies in vielen Fällen nicht als Erinnerungsverlust interpretiert werden, sondern als ein aus den verschiedenen Gründen diskutierten und möglichen Aspekten eines blockierten Zugangs bzw. Schwierigkeiten bei der Dekodierung. An dieser Stelle soll hierbei auf die Möglichkeit und Chance der Grundlagenforschung verwiesen werden, um in diesem speziellen Kontext die entsprechenden Kompetenzen und Beeinträchtigungen bei lernbehinderten Personen noch genauer zu erforschen.

Zudem wäre einerseits zu bedenken, den Zeitraum zwischen Darbietung des Versuchsmaterials und der ersten Befragung zu vergrößern, da Zeugen häufig erst Wochen oder Monate nach der Beobachtung eines Ereignisses von Polizei oder Richter erstmalig vernommen werden. Andererseits muss hierbei die schon mehrfach erwähnte „Kosten-Nutzen-Rechnung“ mehrfacher Befragungen bzw. deren zeitlicher Terminierung sorgsam bedacht werden. So finden sich neben den Befunden der vorliegenden Arbeit etwa einige Belege dafür, dass nicht die reine Anzahl von verbalen Wiederholungen, sondern der Zeitpunkt der Wiederholungen von entscheidender Bedeutung für das mittel- und langfristige Behalten von (bedeutsamen) Erfahrungen ist. Je schneller die erste verbale Wiederholung und damit auch Befragung erfolgt, desto besser ist die langfristige Behaltensleistung (Fivush & Schwarzmüller, 1998; Poole & White, 1995; Shrimpton et al., 1998; Warren & Lane, 1995; zit. nach GREUEL, 2001). Demnach führen die Verbalisation und Wiederholung von Erlebniserinnerungen in erster Linie in der Enkodierungsphase zu einer Verbesserung der Behaltensleistung.

tung. Aufgrund der bisherigen Befunde und auch denen der vorliegenden Arbeit erweist sich insgesamt eine unmittelbare Befragung als vorteilhaft, wenn das Interview zu einem späteren Zeitpunkt wiederholt wird. Dies spräche auf der anderen Seite im Sinne einer (theoretisch) möglichst optimalen, genauen und korrekten Erinnerungsleistung im forensischen Kontext wiederum sehr dafür, einen ersten Befragungszeitpunkt T1 immer so früh wie möglich zu setzen. Andererseits sprechen die Befunde der Ergebnisse der Forschung zu Reminiszenz- und Hypermnesie-Effekten auch dafür, dass unter bestimmten (Befragungs-) Umständen auch einige Details oder Aspekte erst zu späteren Zeitpunkten zugänglich, abrufbar oder erinnerbar sind. Dies hängt natürlich individuell unter anderem von der Qualität der vorangegangenen Befragungen, dem Einfluss nachträglicher Informationsquellen und einigen anderen sozialen, kognitiven und auch motivationalen Faktoren ab. Hier besteht vor allem bei der Population der intellektuell beeinträchtigten Menschen noch immenser Forschungsbedarf im angewandten wie auch im Grundlagenbereich, um diesbezüglich genauere Erkenntnisse zu gewinnen.

Darüber hinaus könnten auch zusätzliche, weitere Erhebungszeitpunkte (T3, T4... oder auch NT3, NT4...) weitere Einblicke und Aufschluss hinsichtlich der Veränderung von Gedächtnis- und Erinnerungsleistungen über die Zeit hinweg geben. Dies könnte auch weitere Aufschlüsse darüber liefern, ob sich die Ergebnisse der vorliegenden Arbeit in ähnlicher oder anderer Tendenz fortsetzen.

Hinsichtlich interrogativer Suggestibilität, allem was die Empfänglichkeit für negatives Feedback betrifft sowie im Rahmen der Entwicklung angemessener Interviewtechniken im Umgang bzw. bei der Befragung intellektuell beeinträchtigter Menschen, muss beachtet werden, dass gerade solche Personengruppen in der Regel eher weniger mit Befragungssituationen und entsprechenden Umständen vertraut sind. So haben etwa Verschiedene Untersuchungen (z.B. CASSEL, ROEBERS & BJORKLUND, 1996; POOLE & WHITE, 1991) gezeigt, dass gerade Kinder dazu neigen, ihre erste – möglicherweise richtige - Antwort zu ändern und eine andere Antwort zu geben, wenn ihnen in kurzem zeitlichem Abstand wiederholt dieselbe (geschlossene) Frage gestellt wird. Dieser Effekt tritt dagegen nicht oder in sehr viel geringerem Maß auf, wenn offene Fragen wiederholt gestellt werden. Dieses Phänomen wird nach Poole & White damit erklärt, dass Befragungen nicht einfach Formen von Wissens- oder Gedächtnistests sind – zumindest nicht explizit – sondern Konversation, die nach gelernten Konversationsregeln geführt wird, auch wenn dies in der forensischen Praxis, aber auch in der wissenschaftlichen Forschung oftmals nicht ganz so ablaufen kann oder soll. Da gerade Kinder – wie auch behinderte Menschen - auf offene Fragen relativ wenig Information produzieren, besteht hier schnell eine starke Tendenz, auf direkte geschlossene Fragen zurückzugreifen. Dabei impliziert das Wiederholen einer Frage oft, dass die erste Antwort

unzulänglich oder falsch war, besonders wenn dies zudem mit einer negativen Rückmeldung verbunden ist. Dies entspricht einer typischen, nicht nur alltäglichen, sondern leider auch in der forensischen oder polizeilichen Vernehmungspraxis üblichen Kommunikationssituation zwischen Erwachsenen und Kindern, bei der Erwachsene häufig Fragen stellen, auf die sie selbst die richtige Antwort kennen oder diese zumindest vermuten und es nun Aufgabe des Kindes ist, diese „richtige“ Antwort zu finden. Eine mögliche Lösung, diese Signalwirkung zu reduzieren, besteht darin, eine Fragewiederholung damit zu begründen, dass der Interviewer etwas nicht richtig verstanden hat, um nicht den Eindruck nahe zu legen, das Kind habe einen Fehler gemacht.

Wie schon dargestellt, ist nach gegenwärtiger Befundlage davon auszugehen, dass positive Effekte von Befragungswiederholungen auf die Aussagegenauigkeit, Fehlerhaftigkeit und Suggestionsresistenz vor allem vom Zeitpunkt der Erstbefragung abhängen, d.h. an eine möglichst frühzeitige Erstbefragung, wenn möglich innerhalb weniger Tage, gebunden sind.

Mittlerweile existiert eine Vielzahl von empirischen Studien zur Validitätsprüfung der aussageimmanenten Realkennzeichen (vgl. u.a. DAHLE, 1997; STELLER et al., 1992; KRAHÉ & KUNDROTAS, 1992; STELLER & KÖHNKEN, 1989). Die Untersuchungsergebnisse dieser Studien deuten insgesamt darauf hin, dass die Qualitätsmerkmale im Sinne der 'Undeutsch-Hypothese' grundsätzlich zwischen erlebnisbasierten und nicht-erlebnisbasierten Zeugenaussagen diskriminieren können.

FABIAN et al. (1997) weisen jedoch darauf hin, dass unterschiedliche Aussagemerkmale je nach Besonderheiten der Aussageperson (z.B. Alter, sprachliche Ausdrucksfähigkeit), des inkriminierten Geschehens sowie der Aussageentstehung jeweils differentiellen Aussagewert haben. So kommen auch FIEDLER & SCHMID (1999) kommen bei ihrer Bestandsaufnahme der empirischen Erkenntnisse zu einem ähnlichen Ergebnis. Sie stellen zum einen fest, dass die „Aussagemerkmale auf keinen Fall den Status von nomologischen Gesetzen beanspruchen dürfen“, zum anderen aber, dass diese Realkennzeichen „durchaus“ als „nützliche Indikatoren“ dazu geeignet sind „im Rahmen klar spezifizierter und kritisch geprüfter Modelle einen bedeutsamen statistischen Beitrag zur Wahrheitsfindung zu leisten“ (FIEDLER & SCHMID, 1999, S. 17).

Die vorliegende Validitätsuntersuchung zur kriterienorientierten Aussageanalyse für Aussagen von Jugendlichen mit einer Intelligenzminderung scheint den bisherigen Erkenntnisstand zu stärken. Auch in dieser Untersuchung können einige Realkennzeichen nicht zwischen erlebnisbasierten und nicht-erlebnisbasierten Aussagen trennen. Andere Merkmale dagegen können wiederum überzeugen. Daher scheint es sich zu bestätigen, dass bei unterschiedlichen Zeugengruppen (Kinder, Erwachsene, Jugendliche mit Intelligenzminderung) unter-

schiedlichen Qualitätsmerkmalen eine besondere Bedeutung zukommt. Hinsichtlich der Einschätzung des Erlebnishintergrundes einer Aussage lässt sich schlussfolgernd somit anmerken, dass erlebnisfundierte Aussagen sich eben nicht allein durch eine erhöhte Anzahl oder Ausprägung einzelner Qualitätsmerkmale, sondern vielmehr durch spezifische Merkmalskonfigurationen und –kombinationen charakterisieren lassen. Somit wäre dann möglicherweise zu erwarten, dass perzeptuelle *und* reflexive Inhalte gleichermaßen in qualifizierter Ausprägung vorliegen. Auf diesem Hintergrund scheint die in Anlehnung an Larsen (1998, zit. nach GREUEL, 2001) vorgenommene Differenzierung der 'speziellen Qualitätsmerkmale des Aussageinhalts' in perzeptuelle und reflexive Aussageinhalte inhaltlich sinnvoll und für die Zwecke der aussagepsychologischen Begutachtung nützlich (vgl. auch GREUEL, 2001).

In diesem Zusammenhang ist nochmals darauf hinzuweisen, dass die aussageimmanente Qualitätsanalyse zwar ein sehr zentrales, nicht das alleinige Element einer Glaubhaftigkeitsbegutachtung ist. Vielmehr handelt es sich in der Praxis immer um einen integrativen Prozess, der die drei Konstrukte Aussagetüchtigkeit, Aussagequalität und Aussagevalidität berücksichtigt. Demnach sollte kein Sachverständiger seine Wahrscheinlichkeitsaussage isoliert auf die aussageimmanente Analyse stützen.

Im Rahmen dieser Validierungsuntersuchung nahmen 25 Versuchspersonen an der Erhebung teil. Somit standen innerhalb eines 'within-subject-Designs' über zwei Befragungszeitpunkte insgesamt 100 Aussagen zur statistischen Auswertung zur Verfügung. Mit dieser Stichprobengröße konnten erste Hinweise zwar gegeben werden, es ist jedoch wünschenswert in Folgeuntersuchungen die Stichprobe zahlenmäßig auszubauen.

Zwar wurde im Rahmen der methodischen Vorüberlegungen zu der vorliegenden Arbeit aufgrund des dargestellten Ausschlussprinzips (Vorauswahl) darauf geachtet, nur bestimmte, möglichst untereinander vergleichbare Versuchspersonen (Jugendliche) in die Stichprobe mit einzubeziehen; da aber eine Lernbehinderung bzw. Intelligenzminderung generell sehr unterschiedlich in ihrem Schweregrad und ihrer speziellen Form ausgeprägt sein kann und der begriffliche Umgang wie ausführlich dargestellt ohnehin kritisch ist, wäre es sinnvoll in zukünftigen Arbeiten einen äußerst sorgfältigen Fokus darauf zu legen, homogenere Gruppen mit bestimmten vergleichbaren Störungsbildern (z.B. leichte Intelligenzminderung, mittelgradige Intelligenzminderung) zu bilden und diese im Vergleich mit entsprechenden Kontrollgruppen zu untersuchen, um so aussagekräftigere Aussagen machen zu können.

Daneben ist ein umfangreiches Training der oder des Raters für eine Validitätsuntersuchung der Glaubhaftigkeitsmerkmale sowie darüber hinausgehendes Wissen und Schulungen im

Umgang mit bzw. den sozialen, emotionalen und aussagepsychologischen Besonderheiten von lernbehinderten Personen unumgänglich, um eine Generalisierung der Ergebnisse möglich zu machen.

Nach FIEDLER & SCHMID (1999) gibt es, neben den in der vorliegenden Untersuchung gemachten nicht-erlebnisbasierten Aussagen, bei denen es sich ja um intentionale Falschaussagen handelt, noch eine Reihe von anderen psychologischen Prozessen, die für das Abweichen von der Wahrheit verantwortlich sein können. Zu diesen Prozessen zählen unter anderem eine fehlerhafte Wahrnehmung, Vergessen, Gedächtnisverzerrungen z.B. durch Quellenverwechslung sowie bewusste oder unbewusste Suggestion. Diese Studie kann daher nicht den Anspruch erheben, für alle möglichen Arten einer Falschaussage Erkenntnisse liefern zu können, versuchte allerdings dennoch bestimmte Schwerpunkte zu setzen. Untersuchungen, die über intentionale Falschaussagen hinausgehen wären wünschenswert. Im Zusammenhang mit bewusst suggestiver Manipulation der Versuchspersonen sind jedoch ethische Bedenken angebracht.

Unter Bezugnahme auf GREUEL (2001) seien im Zusammenhang mit verschiedenen gedächtnispsychologischen Modellen und Befunden hier nur ansatzweise einige kritische Implikationen für die weitere Entwicklung einer ganzheitlichen Psychologie der Aussage angemerkt:

- Für die psychologische Auseinandersetzung mit Aussagen von „Opferzeugen“ sei eine stärkere Fokussierung auf episodische und autobiographische Gedächtnisprozesse zu fordern.
- Da im forensischen Kontext relevante Aussagen von „Opferzeugen“ eine distinkte Form ereignisspezifischer Erinnerungen ausmachen, die aufgrund ihrer Selbst-Bedeutsamkeit, Neuartigkeit, Folgenschwere und emotionalen Intensität nicht nur zentrale, idiosynkratische Organisationsstrukturen des autobiographischen Gedächtnisses konstituieren, sondern gleichzeitig auch selbst hoch memorierbare Einzelepisoden repräsentieren, kann diese distinkte Form ereignisspezifischer Erinnerungen im Sinne von ‘temporal landmarks’ experimentell nur sehr schwer induziert oder simuliert werden, so dass eine Übertragbarkeit traditioneller gedächtnispsychologischer Studien zumindest eingeschränkt und kritisch zu betrachten ist.
- Im Übrigen ist in jedem Fall eine Differenzierung zwischen dem theoretischen Konstrukt „Erinnerung“ und dem operational bestimmbaren „Erinnerungsbericht“ vonnöten, die wie schon mehrfach erwähnt eine über wahrnehmungs- und gedächtnispsychologische Prozesse hinausgehende Auseinandersetzung mit den sozialen bzw. sozialpsychologischen und auch motivationalen Kontextbedingungen von Aussagen beinhaltet.

Gerade Zeugen mit einer Intelligenzminderung stellen für die forensische Praxis, d.h. für Polizei, Gericht und Sachverständige eine besondere Herausforderung dar, da diese Menschen einem erhöhten Risiko, Opfer einer sexuellen Missbrauchs zu werden, ausgesetzt sind und außerdem eine sehr heterogene Zeugengruppe darstellen.

Daher ist es von besonderem Interesse, weitere Forschungsarbeiten zur Zeugenkompetenz durchzuführen, um detaillierte Erkenntnisse zu gewinnen, welche konkreten kognitiven Voraussetzungen in bestimmten Bereichen und unter Einfluss unterschiedlicher Faktoren bei dieser Zeugengruppe gegeben sind. Die Ergebnisse dieser Untersuchung geben einen Hinweis darauf, dass Jugendliche mit einer Intelligenzminderung - entgegen verschiedener Positionen - sehr wohl in der Lage sind eine gerichtlich verwertbare Aussage zu liefern. Die Aussagetüchtigkeit dieser Stichprobe ist unbestritten. Sicherlich bedarf es noch weiterer Untersuchungen um allgemeingültige Aussagen darüber treffen zu können, inwieweit Menschen mit einer Intelligenzminderung generell aussagetüchtig sind. Ein interessanter und hilfreicher Aspekt scheint in diesem Zusammenhang das Entwicklungsalter zu sein. Für Polizei, Richter und Sachverständige, die Zeugen zu Straftatbeständen vernehmen, wäre es in diesem Zusammenhang wünschenswert, dass ein Bewusstsein bei allen Beteiligten entsteht, dass Menschen mit einer Intelligenzminderung spezifische Unterstützungen benötigen.

So können insbesondere spezifische Interviewtechniken in einem erhöhten Maße dazu beitragen, dass die Zeugen adäquat in ihrer Aussage unterstützt werden und ihre individuellen Stärken gefördert werden.

Eine Auseinandersetzung aller Beteiligten mit dem Thema „Zeugen mit Intelligenzminderung“ in Form von Fortbildungen oder Seminaren zur Weiterbildung könnte die Arbeit mit dieser Zeugengruppe professionalisieren.

ZUSAMMENFASSUNG

Im Rahmen der Psychologie der Zeugenaussage und Zeugenkompetenz sowie der Begutachtung im forensischen Kontext stellen zum einen die Analyse der Korrektheit und Genauigkeit wie auch verschiedener Fehlerarten durch Einflüsse aus den unterschiedlichsten Bereichen sowie die Fähigkeit von Zeugen zu unterscheiden, welche Informationen verschiedenen Arten von Quellen zuzuordnen sind (Quellenidentifikation), wichtige und interessante Fragen und Forschungszweige dar. Zum anderen ist auch die Validierung von aussageimmanenten Qualitätsmerkmalen der kriterienorientierten Aussageanalyse von besonderer Relevanz. Defizitär ist die diesbezügliche empirische Befundlage allerdings im Bereich der spezifischen Personengruppe lernbehinderter Jugendlicher bzw. Personen mit einer Intelligenzminderung. In der vorliegenden Arbeit wurden deshalb unter jeweils zusätzlicher Betrachtung eines längsschnittlichen Aspekts (wiederholter Befragungen) in zwei Teiluntersuchungen die oben genannten Fragestellungen an Jugendlichen mit kognitiven Beeinträchtigungen untersucht.

Dabei stellt das erhöhte Risiko dieser Menschen, Opfer von Straftaten gegen das sexuelle Selbstbestimmungsrecht (sexuellen Missbrauchs oder sexueller Gewalt) zu werden, den angewandeten Hintergrund der vorliegenden Arbeit.

Eine kurze und konzentrierte Zusammenfassung der doch sehr umfangreichen, komplexen und differenzierten Befunde ist an dieser Stelle nur schwer möglich. Deshalb sei (und muss) der interessierte Leser auf die entsprechenden Kapitel verwiesen (werden). Dennoch sollen in etwa die charakteristischen Eckpunkte und Teile der Ergebnisse zusammengefasst werden.

Dazu nahmen als Versuchspersonen 44 (im Rahmen der Studie zur Erinnerungs- und Gedächtnisleistung) bzw. 25 (im Rahmen der Validierungsuntersuchung) Jugendliche mit einer Intelligenzminderung (Lernbehinderung) im Durchschnittsalter von jeweils ca. 18 Jahren an den Untersuchungen teil.

Im Rahmen der **Untersuchung zur Gedächtnisleistung und Quellenidentifikation** wurden den Jugendlichen Dia-Serien eines (inszenierten) Verkehrsunfalls präsentiert. Dabei sah eine Hälfte der Versuchspersonen (Bedingung 'external-external') in direkter Aufeinanderfolge zwei semantisch ähnliche Dia-Serien (Unfall 'PKW-PKW' bzw. 'PKW-Fahrrad'), die andere Hälfte ebenfalls die Dia-Serie 'PKW-PKW', musste sich jedoch in Analogie zur Dia-Serie 'PKW-Fahrrad' unter entsprechender Instruktion und Vorgaben eine selbst erfundene Unfall-Serie 'PKW-Fahrrad' ausdenken und deren Ablauf und beteiligte Personen anhand

eines auszufüllenden Bogen beschreiben und schriftlich niederlegen (Bedingung 'external-internal').

Wiederum jeweils die Hälfte Versuchspersonen der beiden Versuchsgruppen wurde in halbstrukturierten Interviews (freier Bericht mit anschließendem spezifischem Nachfragen; Trichtertechnik) entweder zweimal zu den Dia- bzw. Unfall-Serien befragt (erstmal eine Woche und zum zweiten Mal ca. acht Wochen nach Präsentation des Versuchsmaterials) oder nur einmal ebenfalls nach acht Wochen.

Abhängige Variablen waren dabei jeweils die Anzahl korrekt berichteter Details (im freien Bericht wie auf Nachfrage) wie verschiedene Arten von berichteten bzw. begangenen Fehlern (Gesamtfehlerzahl, allgemeine Erinnerungsfehler, Quellenverwechslungen, schematypische Intrusionen und Konfabulationen).

Insgesamt ergaben sich sehr differenzierte Befunde zwischen den Versuchsgruppen, wie auch hinsichtlich der Veränderung der Gedächtnisleistungen über die Zeit und zu den einzelnen Befragungszeitpunkten.

Betrachtet man die Befragungszeitpunkte einzeln, berichten die beiden Versuchsgruppen zu keinem der Befragungszeitpunkte insgesamt wesentlich mehr oder weniger Details korrekt über beide Serien hinweg. Deutliche Unterschiede ergeben sich jedoch zu allen Befragungszeitpunkten hinsichtlich der Dia-Serie 'PKW-PKW', derart, dass die Jugendlichen der Bedingung 'external-internal' jeweils deutlich mehr korrekte Details berichteten. Hinsichtlich der Unfall-Serie 'PKW-Fahrrad' ergaben sich allerdings nur bei einer zweiten Befragung nach acht Wochen deutliche Unterschiede im Bereich der insgesamt korrekt berichteten Details zwischen den beiden Versuchsbedingungen.

Hinsichtlich der insgesamt begangenen Fehler ergibt sich für alle Befragungszeitpunkte und über alle Serien hinweg (zusammen wie auch einzeln betrachtet) ein einheitliches Bild. Die Versuchspersonen der Bedingung 'external-internal' produzieren hier jeweils deutlich mehr Fehler als die der Bedingung 'external-external'. Quellenverwechslungen scheinen zu allen drei Befragungszeitpunkten lediglich innerhalb der Serien von entscheidender Bedeutung zu sein und hier insbesondere hinsichtlich der Unfall-Serie 'PKW-Fahrrad'.

Über die Zeit hinweg betrachtet, also zwischen den beiden Befragungszeitpunkten T1 und T2 (nach einer bzw. acht Wochen) scheinen sich Vergessensprozesse in Form von weniger korrekt berichteten Details lediglich eher für die Gruppe 'external-internal' negativ auszuwirken. In der externalen Quellenüberwachungsgruppe sind die diesbezüglichen Befunde offenbar konstant über die Zeit.

Hinsichtlich der insgesamt sowie im Bereich der allgemeinen Erinnerungsfehler begangenen Fehler konnten erwartungsgemäß bei einem zweiten Befragungszeitpunkt nach acht Wochen in beiden Versuchsgruppen deutlich höhere Fehlerquoten gefunden werden.

Wesentliche Veränderungen über die Zeit im Bereich von Quellenverwechslungen konnten allerdings in beiden Versuchsgruppen nicht festgestellt werden. insgesamt wurden ohnehin nur sehr wenige Quellenverwechslungen begangen.

Schematypische Intrusionen sowie Konfabulationen scheinen über die Zeit hinweg zumindest in der Gruppe 'external-external' ebenso recht stabil zu sein, da hier keine deutlichen Veränderungen gefunden wurden. Etwas differenziertere Befunde gelten jedoch für die Versuchsgruppe 'external-internal'.

Ebenfalls recht differenziert stellen sich die Befunde bezüglich der Vergleiche dar bezogen auf die Frage, ob bzw. welchen Unterschied es zum einen macht, ob eine erste Befragung nach einer oder nach acht Wochen stattfindet, zum anderen ob und welche sich Unterschiede bei einer ersten gegenüber einer zweiten Befragung nach acht Wochen ergeben.

Generell kann jedoch festgehalten werden, dass es im Bereich der insgesamt begangenen Fehler wie auch der allgemeinen Erinnerungsfehler für die Jugendlichen der Versuchsbedingung 'external-internal', wenn diese also zwischen einer gesehenen Dia-Serie und einer intern selbst generierten Quellen differenzieren mussten, bei einer ersten Befragung nach acht Wochen gegenüber sowohl einer ersten Befragung nach einer Woche wie auch einer zweiten nach ebenfalls acht Wochen, zu jeweils wesentlich höheren Fehlerraten kommt. Hinsichtlich einer Minimierung der Fehlerquote sprechen diese Befunde zumindest in jedem Fall gegen eine erst sehr späte Erstbefragung von Zeugen. In der Gruppe der Bedingung 'external-external', wenn also zwei Dia-Serien gesehen wurden, ergaben sich diesbezüglich jedoch für keinen der Vergleiche (T1-NT2 und T2-NT2) wesentliche Unterschiede.

Wiederum wurden beim Vergleich der genannten Befragungszeitpunkte für keine der Versuchsgruppen Unterschiede im Bereich von Quellenverwechslungen festgestellt. Für lernbehinderte Jugendliche scheint eine zeitliche Dimension (Aspekt) – egal, ob zwischen zwei Befragungen (T1-T2) oder hinsichtlich der eben beschriebenen Vergleiche (T1-NT2 oder T2-NT2) – bezüglich der Gefahr von falschen Quellenzuschreibungen von sehr geringer Relevanz zu sein.

Zudem scheinen konfabulierte Inhalte beim Vergleich einer ersten Befragung (nach jeweils acht Wochen) wie auch beim Vergleich einer ersten Befragung nach einer gegenüber einer zweiten nach acht Wochen für Jugendliche beider Versuchsbedingungen von geringerer Bedeutung zu sein als etwa schematypische Intrusionen, für die sich zu den genannten temporären Vergleichen in den beiden Versuchsbedingungen jeweils etwa analoge Befunde ergaben.

Auch hinsichtlich korrekt berichteter Details ergaben sich für die beiden Versuchsgruppen interessante Befunde, die jedoch einer differenzierten Betrachtung bedürfen.

Im Rahmen der Befunde zu den Haupt- und Interaktionseffekten sind wie schon angedeutet vor allem die Aspekte der insgesamt begangenen Fehler wie auch der allgemeinen Erinnerungsfehler zu nennen.

Ein bedeutender Effekt des Befragungszeitpunktes konnte neben den gerade genannten lediglich noch für den Bereich der Konfabulationen gefunden werden. Wechselseitige Abhängigkeiten der beiden Faktoren 'Gruppe' und 'Befragungszeitpunkt' in Form von statistisch signifikanten Interaktionen fanden sich wiederum nur für den Bereich der Gesamtfehlerzahl wie den der allgemeinen Erinnerungsfehler.

Im Rahmen der **Validierungsuntersuchung**, die nach Kenntnis des Verfassers der vorliegenden Arbeit, die bisher einzige und erste ist, die sich mit der speziellen Personengruppe der lernbehinderten Jugendlichen befasste, sollten die Versuchspersonen nach Fremdinstruktion (Blindversuch) jeweils ein emotional nicht neutrales autobiographisches Erlebnis wie auch ein thematisch ähnliches bzw. bestimmte in Simulationsstudien dieser Art übliche Grundvariablen (Kriterien) der emotionalen Involviertheit erfüllendes erfundenes (nicht erlebnis-basiertes) Ereignis schildern.

Nach ca. sechs Wochen fand ein zweiter Interviewtermin statt, an dem ebenfalls beide Geschichten berichtet werden sollten. Somit lagen also von jeder Versuchsperson zwei Aussagen aus zwei Befragungsterminen vor, insgesamt also 100 Aussagen, 50 erlebnisfundierte und 50 ohne Erlebnishintergrund.

Abhängige Variablen waren die unter anderem in einem Urteil des BGH zur Begutachtung von Zeugenaussagen festgesetzten Glaubhaftigkeitsmerkmale (auch Realkennzeichen oder Qualitätsmerkmale). So war es Ziel und Aufgabe der vorliegenden Teiluntersuchung festzustellen, inwieweit bzw. welche der Qualitätsmerkmale zwischen erlebnisbasierten und nicht erlebnisbasierten Aussagen differenzieren können und wie sich die aussageimmanenten Inhalte über die Zeit hinweg zu einem zweiten Befragungszeitpunkt verändern.

Auf dem Hintergrund der Hypothese, dass sich in erlebnisfundierten Aussagen deutlich höhere Merkmalsausprägungen der einzelnen Aussagemerkmale finden lassen als in solchen ohne Erlebnishintergrund und der allgemeinen Annahme, dass die Merkmalsausprägungen über die Zeit hinweg (von T1 nach T2) abnehmen, fanden sich für die meisten und wichtigsten Glaubhaftigkeitsmerkmale hypothesenkonforme Ergebnisse:

So fanden sich etwa sowohl zu T1 als auch zu T2 in erlebnis-basierten Aussagen für folgende Merkmale deutlich höhere Merkmalsausprägungen auf als in nicht erlebnis-basierten, welche somit sehr gute, vom Zeitpunkt unabhängige Diskriminatoren darstellen:

- Logische Konsistenz
- Widerspruchsfreie, ungeordnete Darstellung
- Kontextuelle Einbettung
- Schilderung nonverbaler Aktionen
- Nebensächliche Details
- Schilderung eigener psychischer oder physischer Vorgänge
- Einwände gegen die Richtigkeit der eigenen Aussage
- Deliktspezifische Aussageelemente

Für keines der Glaubhaftigkeitskriterien war zu keinem der beiden Befragungszeitpunkte in den erfundenen Geschichten eine wesentlich oder statistisch bedeutsame höhere Merkmalsausprägung zu beobachten als in den wahren.

Ein deutlicher Abfall in der Merkmalsausprägung von T1 nach T2 konnte für beide Geschichten lediglich hinsichtlich der 'Schilderung nonverbaler Aktionen' sowie der 'Schilderung eigener psychischer oder physischer Vorgänge' nachgewiesen werden.

Für erlebnis-basierte Aussagen fand sich dieser Rückgang in der Merkmalsausprägung hinsichtlich 'nebensächlicher Details', 'spontaner Verbesserungen der eigenen Aussage' sowie 'Selbstbelastungen'.

Für nicht erlebnis-basierte Aussagen waren bezüglich der 'Wiedergabe von Gesprächen', der 'Schilderung von Komplikationen im Handlungsverlauf' sowie der 'Schilderung 'ausgefallener Details' die Mittelwerte zum zweiten Befragungszeitpunkt T2 geringer als zum ersten Befragungszeitpunkt T1.

Keines der Realkennzeichen wies weder für wahre noch für erfundene Geschichten - wenn überhaupt - einen wesentlichen Anstieg in der Merkmalsausprägung von T1 nach T2 auf.

Folgende Aussagemerkmale trennen zu beiden Befragungszeitpunkten T1 und T2 zwischen Aussagen mit Erlebnishintergrund und Aussagen ohne Erlebnishintergrund:

- Logische Konsistenz
- Ungeordnete, widerspruchsfreie Darstellung
- Quantitativer Detailreichtum
- Kontextuelle Einbettung
- Schilderung nonverbaler Aktionen
- Nebensächliche Details
- Schilderung eigener psychischer oder physischer Vorgänge
- Einwände gegen die Richtigkeit der eigenen Aussage
- Deliktspezifische Aussageelemente

Bei Betrachtung des ersten Befragungszeitpunktes T1 für sich ergibt sich ein deutlicher Unterschied zwischen wahren und erfundenen Geschichten nur hinsichtlich 'spontaner Verbesserungen der eigenen Aussage'.

Allein zum zweiten Befragungszeitpunkt T2 trennen die Realkennzeichen 'Wiedergabe von Gesprächen', 'Schilderung von Komplikationen im Handlungsverlauf' sowie die 'Schilderung ausgefallener Details'.

Betrachtet man die einzelnen *Merkmalskategorien von Qualitätsmerkmalen*, so ergaben sich zusammengefasst folgende Befunde:

Bezüglich der ersten und diagnostisch sehr bedeutsamen Merkmalskategorie 'Allgemeine Merkmale' konnten zu beiden Befragungszeitpunkten T1 und T2 in den erlebnisfundierten Aussagen deutlich höhere Ausprägungen festgestellt werden als in den nicht erlebnisfundierten. Allerdings finden hier weder für wahre noch erfundene Geschichten bedeutsame Veränderungen über die Zeit hinweg von T1 nach T2 statt.

Für den Bereich der 'Speziellen Inhalte' fanden sich zum einen deutlich höhere Merkmalsausprägungen in den wahren Geschichten gegenüber den erfundenen (sowohl zu T1 als auch zu T2), zum anderen zum zweiten Befragungszeitpunkt deutlich geringere als zum ersten Befragungszeitpunkt (sowohl für wahre als auch für erfundene Geschichten).

Exakt analog zu den 'Speziellen Inhalten' gestalten sich die Befunde hinsichtlich der 'Inhaltlichen Besonderheiten'.

Innerhalb des Merkmalskomplexes 'Motivationsbezogene Inhalte' zeigte sich lediglich zum ersten Befragungszeitpunkt T1 ein deutlicher höherer Wert in der Merkmalsausprägung in den erlebnis-basierten Aussagen gegenüber denjenigen ohne Erlebnishintergrund. Wie auch hinsichtlich der 'Allgemeinen Merkmale' verändern sich die Merkmalsausprägungen weder für wahre noch für erfundene Geschichten nicht wesentlich vom ersten zum zweiten Befragungszeitpunkt.

Für keine der Merkmalskategorien konnte eine wechselseitige Abhängigkeit zwischen Erlebnisbasis und Befragungszeitpunkt (Interaktion) beobachtet werden.

Analoge Befunde ergeben sich aus den Ergebnissen der zusätzlich durchgeführten Diskriminanzanalyse.

Zu beiden Befragungszeitpunkten stellten sich hierbei vor allem die Aussagemerkmale aus dem Bereich der 'Allgemeinen Merkmale' als bedeutsamste Diskriminationskriterien heraus, insbesondere das Merkmal 'Logische Konsistenz'.

Als ebenso diskriminatorisch sehr bedeutsam erwiesen sich vor allem die 'Kontextuelle Einbettung' von Aussageelementen sowie die 'Schilderung nonverbaler Aktionen' (spezielle Inhalte) - sowohl zum ersten als auch zum zweiten Befragungszeitpunkt.

Aus dem Bereich der 'Inhaltlichen Besonderheiten' scheint der 'Schilderung eigener psychischer und physischer Vorgänge' - ebenfalls zu beiden Befragungszeitpunkten - ein hoher diskriminatorischer Wert zu zukommen.

Ingesamt kann abschließend festgehalten werden, dass die vorliegende Untersuchung erste Hinweise für die Glaubhaftigkeitsbegutachtung bei Zeugen mit Intelligenzminderung im Rahmen der Analyse der Aussagequalität geben kann. Die Ergebnisse deuten darauf hin, dass Jugendliche mit einer Intelligenzminderung glaubhaft ein erlebnisbasiertes Ereignis schildern können. Zwar trennen einige der Merkmale auch nicht zwischen wahren und erfundenen Aussagen, jedoch finden sich letztendlich in den bedeutsamen Kriterien vor allem aus dem Bereich der 'Allgemeinen Merkmale' wie auch der 'Speziellen Inhalte' und besonders was die 'Schilderung eigener psychischer Vorgänge' angeht, in Schilderungen ohne Erlebnishintergrund doch deutlich geringere Merkmalsausprägungen.

So kann insgesamt (auch zu einem zweiten Befragungszeitpunkt in spezieller Weise mit etwas anderen Gewichtungen) die merkmalsorientierte Aussageanalyse als Teil der sogenannten 'Statement Validity Analysis' (SVA) wertvolle Beiträge bei Begutachtung von Zeugen mit intellektuellen Beeinträchtigungen liefern bzw. als geeignetes Instrument angesehen werden.

Die Befunde der beiden Teiluntersuchungen der Arbeit werden auf Grundlage kognitions-, entwicklungs-, sozial- und motivationspsychologischer Erkenntnisse und speziellen Anforderungen an adäquate Befragungstechniken im Umgang mit der spezifischen Personengruppe der Lernbehinderten und deren besonderen Hintergründen, Defiziten und Anforderungen in Bezug auf ihre Implikationen für die forensische und diagnostische Praxis diskutiert.

SCHLÜSSELBEGRIFFE:

Zeugenaussagen - Gedächtnis - Jugendliche - Intelligenzminderung - Lernbehinderung -
Zeugenkompetenz - Realitätsüberwachung - Quellenkontrolle - Quellendifferenzierung -
Glaubhaftigkeitsdiagnostik - kriterienorientierte Aussageanalyse - wiederholte Befragung

LITERATURVERZEICHNIS

- AGNEW, S. E. & POWELL, M. B. (2004): The Effect of Intellectual Disability on Children's Recall of an Event across Different Question Types. *Law and Human Behavior*, 28 (3), 273-294.
- AKEHURST, L., KÖHNKEN, G. & HÖFER, E. (2001): Content Credibility Of Accounts Derived From Live And Video Presentations. *Legal And Criminological Psychology*, 6, pp. 65-83.
- ALLWOOD, C. M., JONSSON, A.-C. & GRANHAG, P. A. (2004): The Effects of Source and Type of Feedback on Child Witnesses' Metamemory Accuracy. *Applied Cognitive Psychology*, ??, in press.
- ANSON, D. A., GOLDING, S. L. & GULLY, K. J. (1993): Child Sexual Abuse Allegations: Reliability Of Criteria-Based Content Analysis. *Law And Human Behavior*, 17 (3), pp. 331-340
- ARNTZEN, F. (1993): *Psychologie der Zeugenaussage. System der Glaubwürdigkeitsmerkmale*. München: Beck.
- ATWELL, J. A., CONNERS, F. A. & MERRILL, E. C. (2003): Implicit and Explicit Learning in Young Adults with Mental Retardation. *American Journal on Mental Retardation*, 108 (1), 56-68.
- BADDELEY, A. D. (1986): *Working Memory*. New York: Oxford University Press.
- BALLOFF, R. (2000): Das Urteil des BGH vom 30. Juli 1999 zur Frage der wissenschaftlichen Anforderungen an aussagepsychologische Begutachtungen (Glaubhaftigkeitsgutachten) und die Folgen für die Sachverständigentätigkeit. In: *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 49 (4), pp. 261-274.
- BALOGH, R. et al. (2001): Sexual Abuse in Children and Adolescents with Intellectual Disability. *Journal of Intellectual Disability Research*, 45 (3), 194-201.
- BATCHELDER, W. H. & RIEFER, D. M. (1990): Multinomial Processing Models of Source Monitoring. *Psychological Review*, 97, 548-564.

- BAYEN, U. J. & MURNANE, K. (1996): Aging and the use of perceptual and temporal information in Source Memory tasks. *Psychology and Aging*, 11, 293-303.
- BEAIL, N. (2002). Interrogative Suggestibility, Memory and Intellectual Disability. *Journal of Applied Research in Intellectual Disability*, 15, 129-137.
- BECKER, M. (1995). *Sexuelle Gewalt gegen Mädchen mit geistiger Behinderung. Daten und Hintergründe*. Heidelberg: Edition Schindele.
- BEKERIAN, D. A. & BOWERS, J. M. (1983): Eyewitness Testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 139-145.
- BELLI, R. F., WINDSCHITL, P. D., MCCARTHY, T. T. & WINFREY, S. E. (1992): Detecting Memory Impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 356-367.
- BENDER, R. & NACK, A. (1995A): *Tatsachenfeststellung vor Gericht. Band 1: Glaubwürdigkeits- und Beweislehre*. München: Beck.
- BENDER, R. & NACK, A. (1995B): *Tatsachenfeststellung vor Gericht. Band 2: Vernehmungslehre*. München: Beck.
- BERGER, O. (2002): *Ein Experiment zur Validierung der kriterienorientierten Aussageanalyse bei Kindern. Untersuchungen der aussageimmanenten Realkennzeichen im Längsschnitt*. Unveröffentlichte Diplomarbeit: Universität Regensburg.
- BERNTSEN, D. (2001): Involuntary Memories of Emotional Events: Do Memories of Traumas and Extremely happy Events Differ? *Applied Cognitive Psychology*, 15, 135-158.
- BERNTSEN, D. (2002): Tunnel Memories for Autobiographical Events: Central Details are remembered more frequently from shocking than from happy experiences. *Memory & Cognition*, 30 (7), 1010-1020.
- BLEIDICK, U. et al. (1977). *Einführung in die Behindertenpädagogik* (Band 1 und Band 2). Stuttgart: Kohlhammer.

- BLUCK, S. & LI, K. Z. H. (2001): Predicting Memory Completeness and Accuracy: Emotion and Exposure in Repeated Autobiographical Recall. *Applied Cognitive Psychology*, 15, 145-158.
- BORKOWSKI, J. G. & BÜCHEL, F. P. (1983): Learning and Memory Strategies in the Mentally Retarded. In: Pressley, M. & Levin, J. R. (Eds.): *Cognitive Strategy Research – Psychological Foundations*. pp. 103-128. New York: Springer.
- BORUM, R. & OTTO, R. (2000): Advances in Forensic Assessment and Treatment: An Overview and Introduction to the Special Issue. *Law and Human Behavior*, 24 (1), 1-7.
- BOWLING, A. C. & MACKENZIE, B. D. (1996): The Relationship between Speed of Information Processing and Cognitive Ability. *Personality and Individual Differences*, 20 (6), 775-800.
- BRIGHAM, J. C. (1990): Target Person Distinctiveness and Attractiveness as Moderator Variables in the Confidence-Accuracy Relationship in Eyewitness Identifications. *Basic and Applied Social Psychology*, 11 (1), 101-115.
- BROWN, C. L. & GEISELMAN, E. R. (1990): Eyewitness Testimony of Mentally Retarded: Effect of the Cognitive Interview. *Journal of Police and Criminal Psychology*, 6, 14 –21.
- BROWN, G. D. A. & CHATER, N. (2001): The Chronological Organization of Memory: Common Psychological Foundations for Remembering and Timing. In: Hoerl, C. & McGormack, T. (Eds.): *Time and Memory – Issues in Philosophy and Psychology*, pp.77-110. Oxford: Clarendon Press.
- BROWN, S. C. & CRAIK, F. I. M. (2000): Encoding and Retrieval of Information. In: Tulving, E. & Craik, F. I. M. (Eds.): *The Oxford Handbook of Memory*, pp. 93-106. Oxford: University Press.
- BROXHOLME, S. L. & LINDSAY, W. R. (2003): Development and Preliminary Evaluation of a Questionnaire on Cognitions related to Sex Offending for Use with Individuals who have mild Intellectual Disabilities. *Journal of Intellectual Disability Research*, 47 (6), 472-482.

- BRUCK, M. & CECI, S. (2004): Unveiling four Common Misconceptions. *Forensic Developmental Psychology*, 13 (6), 229-232.
- BRUCK, M., CECI, S. J. & HEMBROOKE, H. (2002): The Nature of Children's True and False Narratives. *Developmental Review*, 22, 520-554.
- BRUCK, M. & MELNYK, L. (2004): Individual Differences in Children's Suggestibility: A Review and Synthesis. *Applied Cognitive Psychology*, 18, 947-996.
- BRYAN, T. & BRYAN, J. (1990): Social Factors in Learning Disabilities: An Overview. In: Swanson, H. L. & Keogh, B. (Eds.): *Learning Disabilities: Theoretical and Research Issues*. Hillsdale: Erlbaum.
- BULL, R. (1995A): Interviewing Children in Legal Contexts. In: Bull, R. & Carson, D.: *Handbook of Psychology in Legal Contexts*, pp. 235 - 246. Chichester: Wiley.
- BULL, R. (1995B): Interviewing People with Communicative Disabilities. In: Bull, R. & Carson, D. (Eds.): *Handbook of Psychology in Legal Contexts*, pp. 247-260. Chichester: Wiley.
- BULL, R. (1995C): Innovative Techniques for the Questioning of Child Witnesses, especially those who are young and those with Learning Disability. In: Zaragoza, M. S., Graham, J. R., Hall, G., Hirschman, R. & Ben-Porath, Y. S.: *Memory and Testimony in the Child Witness*, pp.179 - 194. London: Sage Publications
- BULL, R. (1997): Interviewing Children in the Forensic Context. In: Greuel, L., Fabian, T. & Stadler, M. (Hrsg.): *Psychologie der Zeugenaussage*. pp. 225-230. Weinheim: Psychologie Verlags Union.
- BUNDESGERICHTSHOF (1999): Urteil des BGH vom 30. Juli 1999 – 1 StR 618/98 – LG Ansbach: Wissenschaftliche Anforderungen an aussagepsychologische Begutachtungen (Glaubhaftigkeitsgutachten). In: *Praxis der Rechtspsychologie*, 9 (2), pp. 113-125.
- BURGWYN-BAILES, E., BAKER-WARD, L., GORDON, B. N. & ORNSTEIN, P. A. (2001): Children's Memory for Emergency Medical Treatment after One Year: The Impact of Individual Difference Variables on Recall and Suggestibility. *Applied Cognitive Psychology*, 15, 25-48.

- BURT, C. D. B., KEMP, S. & CONWAY, M. A. (2003): Themes, Events and Episodes in Autobiographical Memory. *Memory and Cognition*, 31 (2), 317-325.
- BUSSE, D. & VOLBERT, R. (1997): Zur Situation kindlicher Zeugen vor Gericht. In: Steller, M. & Volbert, R. (Hrsg.): *Psychologie im Strafverfahren: Ein Handbuch*. pp. 224-246. Bern: Huber.
- CARDONE, D. & DENT, H. (1996): Memory and Interrogative Suggestibility: the effects of modality of information presentation and retrieval conditions upon the suggestibility scores of people with Learning Disabilities. *Legal Criminal Psychology*, 1, 165 – 177.
- CASSEL, W., ROEBERS, C. & BJORKLUND, D. (1996): *Developmental patterns of Eyewitness responses to repeated and increasingly suggestive questions*. Journal of Experimental Child Psychology, 61, 116-133.
- CECI, S. J. & BRUCK, M. (1995): Suggestibility of the Child Witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439.
- CHAMBERLAIN, A., RAUH, J. & PASSER, A. (1984). Issues in fertility control for Mentally Retarded Female Adolescents. *Pediatrics*, 73(4), 445-450.
- CHAMBERS, K. L. & ZARAGOZA, M. S. (2001): Intended and Unintended Effects of explicit warnings on Eyewitness Suggestibility: Evidence from Source Identification Tests. *Memory & Cognition*, 29 (8), 1120-1129.
- CHANDLER, C. C., GARGANO, G. J. & HOLT, B. (2001): Witnessing Postevents does not change Memory Traces, but can affect their Retrieval. *Applied Cognitive Psychology*, 15, 3-22.
- CHRISTIANSON, S. A. & HUBINETTE, B. (1993). Hands Up ! A study of Witness emotional Reactions and Memories associated with bank robberies. *Applied Cognitive Psychology*, 7, 365-379.
- CHUN, M. M. & JIANG, Y. (2003): Implicit, Long-Term Spatial Contextual Memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29 (2), 224-234.

- CLARE, I. (2001): Witnesses with Learning Disabilities. *British Journal of Learning Disabilities*, 29, 79-80.
- CLARE, I. C. H. & GUDJONSSON, G. H. (1993): Interrogative Suggestibility, Confabulation and Acquiescence in People with Mild Learning Disabilities (Mental Handicap): Implications for Reliability during Police Interrogations. *British Journal of Clinical Psychology*, 32, 295-301.
- COHEN, G. (1996): *Memory in the Real World*. Hove: Psychology Press.
- COHEN, N. J. & FAULKNER, D. (1988). Life span changes in Autobiographical Memory. In: Gruneberg, M., Morris, P. & Sykes, R. N. (Eds.): *Practical Aspects of Memory: Current Research and Issues*. Chichester: Wiley.
- COLWELL, K., HISCOCK, C. K. & MEMON, A. (2002): Interviewing Techniques and the Assessment of Statement Credibility. *Applied Cognitive Psychology*, 16, 287-300.
- COOKE, P. & DAVIES, G. (2001): Achieving Best Evidence from Witnesses with Learning Disabilities: New Guidance. *British Journal of Learning Disabilities*, 29, 84-87.
- CRAIG, R. A., SCHEIBE, R., RASKIN, D. C., KIRCHER, J., C., DODD, D. H. (1999): Interviewer Questions and Content Analysis of Children's Statements of Sexual Abuse. *Applied Developmental Science*, 3 (2), 77-85.
- CUTLER, B. L. & PENROD, S. D. (1995): *Mistaken Identification: The Eyewitness, Psychology and the Law*. Cambridge: University Press.
- D'ARGEMBEAU, A., COMBLAIN, C. & VAN DER LINDEN, M. (2003): Phenomenal Characteristics of Autobiographical Memories for Positive, Negative and Neutral Events. *Applied Cognitive Psychology*, 17, 281-294.
- DAHLE, K.-P. (1997): Wege zu einem linguistischen „Wahrheitstest“? Perspektiven einer einzelfallexperimentellen Weiterentwicklung der Kriterienorientierten Aussageanalyse für die forensisch-psychologische Glaubwürdigkeitsdiagnostik. *Diagnostica*, 43 (1), 3-26.

- DAVIES, G. M. (1993). Witnessing Events. In: Davies, G. M. & Logie, R. H. (Eds.): *Memory in Everyday Life* (pp. 367-401). Amsterdam: North Holland.
- DENT, H. R. (1986): An experimental study of the Effectiveness of Different Techniques of Questioning Mentally Handicapped Child Witnesses. *British Journal of Clinical Psychology*, 25, 13-17.
- DENT, H. (1992): *The Effects of Age and Intelligence on Eyewitness Ability*. In: Dent, H. & Flin, R.: *Children as Witnesses*, pp. 1 - 14. Chichester: Wiley
- DENT, H. & FLIN, R. (1992): *Children as Witnesses*. Chichester: Wiley
- DILLING, H., MOMBOUR, W., SCHMIDT, M. H. (1993) (Hrsg.): *Internationale Klassifikation psychischer Störungen – ICD-10*. Bern: Verlag Hans Huber.
- DOBSON, M. & MARKHAM, R. (1993): Imagery Ability and Source Monitoring: Implications for Eyewitness Memory. *British Journal of Psychology*, 32, 111-118.
- DRIVDAHL, S. B. & ZARAGOZA, M. S. (2001): The Role of Perceptual Elaboration and Individual Differences in the Creation of False Memories for Suggested Events. *Applied Cognitive Psychology*, 15, 265-281.
- EAKIN, D. K., SCHREIBER, T. A. & SERGENT-MARSHALL, S. (2003): Misinformation Effects in Eyewitness Memory: The presence and absence of Memory Impairment as a function of warning and Misinformation Accessibility. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29 (5), 813-825.
- EISEN, M. L. & LYNN, S. J. (2001): Dissociation, Memory and Suggestibility in Adults and Children. *Applied Cognitive Psychology*, 15, 49-73.
- EMMET, D., CLFFORD, B. R. & GWYER, P. (2003): An Investigation of the Interaction between Cognitive Style and Context Reinstatement on the Memory Performance of Eyewitnesses. *Personality and Individual Differences*, 34, 1495-1508.
- ERICSON, K. & ISAACS, B. (2003): Eyewitness Identification Accuracy: A Comparison of Adults with and those without Intellectual Disabilities. *Mental Retardation*, 41 (3), 161-173.

- FEGERT, J. M., BERGER, C., KLOPFER, U., LEHMKUHL, U. & LEHMKUHL, G. (2001): *Ungang mit sexuellem Missbrauch. Forschungsbericht*. Münster: Votum.
- FERGUSON, S., HASHTROUDI, S. & JOHNSON, M. K. (1992): Age differences in using source-relevant cues. *Psychology and Aging*, 7, 443-452.
- FINKELHOR, D. (1984). *Child Sexual Abuse. New Theory and Research*. New York: The Free Press.
- FIVUSH, R. & SCHWARZMUELLER, A. (1995): Say it once again: Effects of Repeated Questions on Children's Event Recall. *Journal of Traumatic Stress*, 8 (4), 555-580.
- FLIN, R., BULL, R., BOON, J., & KNOX, A. (1992): Children in the Witness Box. In: Dent, H., & Flin, R.: *Children as Witnesses*, pp. 167 - 179, Chichester: Wiley.
- FORD, R. M., KEATING, S. & PATEL, R. (2004): Retrieval-Induced Forgetting: A Developmental Study. *British Journal of Developmental Psychology*, 22, 585-603.
- FREUDENBERG, D. (2001): Qualitätsanforderungen an Gutachten aus Sicht der Staatsanwaltschaft. In: Fegert, J. (Hrsg.): *Begutachtung sexuell missbrauchter Kinder*. pp. 68-82. Neuwied; Kriftel: Luchterhand.
- FROST, P., INGRAHAM, M. & WILSON, B. (2002): Why Misinformation is more likely to be recognised over time: A Source Monitoring Account. *Memory*, 10 (3), 179-185.
- GABBERT, F., MEMON, A. & ALLAN, K. (2003): Memory Conformity: Can Eyewitnesses Influence Each Other's Memories for an Event? *Applied Cognitive Psychology*, 17, 533-543.
- GOLDSMITH, M., KORIAT, A. & PANSKY, A. (2005): Strategic Regulation of Grain Size in Memory Reporting Over Time. *Journal of Memory and Language*, xx, xxx-xxx (in press).
- GORDON, B. N., JENS, K. G., HOLLING, R. & WATSON, T. E. (1994): Remembering Activities Performed versus those Imagined: Implications for Testimony of Children with Mental Retardation. *Journal of Clinical Child Psychology*, 23 (3), 239-248.

- GREENBERG, M. S., WESTCOTT, D. R. & BAILEY, S. E. (1998): When Believing is Seeing: The Effect of Scripts on Eyewitness Memory. *Law and Human Behavior*, 22 (6), 685-694.
- GREUEL, L. (2000): Qualitätsstandards aussagepsychologischer Gutachten zur Glaubhaftigkeit von Zeugenaussagen. In: *Monatszeitschrift für Kriminologie und Strafrechtsreform*, 83 (2), pp. 59-70.
- GREUEL, L. (2001): *Wirklichkeit – Erinnerung – Aussage*. Weinheim: Psychologie Verlags Union, Beltz.
- GREUEL, L., FABIAN, T. & STADLER, M. (1997) (Hrsg.): *Psychologie der Zeugenaussage. Ergebnisse der rechtspsychologischen Forschung*. Weinheim: Psychologie Verlags Union, Beltz.
- GREUEL, L., OFFE, S, FABIAN, A., WETZELS, P., FABIAN, T., OFFE, H, STADLER, M. (1998): *Glaubhaftigkeit der Zeugenaussage*. Weinheim: Psychologie Verlags Union, Beltz.
- GROSSMAN, H. J. (1983). *Classification in Mental Retardation*. Washington, DC: American Association on Mental Retardation.
- GUDJONSSON, G. (1992): *The Psychology of Interrogations, Confessions and Testimony*. Chichester: Wiley.
- GUDJONSSON, G., CLARE, I., RUTTER, S. & PEARCE, J. (1993): Persons at Risk during Interviews in Police Custody: The Identification of Vulnerabilities. *Research Study No. 12 for the Royal Commission On Criminal Justice*. London: HMSO
- GUDJONSSON, G. H. & CLARE, I. (1995): The Relationship between Confabulation and Intellectual Ability, Memory, Interrogative Suggestibility and Acquiescence. *Personality and Individual Differences*, 19 (3), 333-338.
- GUDJONSSON, G. H., MURPHY, G. H. & CLARE, I. C. H. (2000): Assessing the Capacity of People with Intellectual Disabilities to be Witnesses in Court. *Psychological Medicine*, 30, 307-314.

- HALLSTEIN, M. (1993). Sexueller Missbrauch und geistige Behinderung. In: Voss, A. & Hallstein, M. (Hrsg.), *Menschen mit Behinderungen. Berichte, Erfahrungen, Ideen zur Präventionsarbeit*. Ruhnmark: Donna Vita.
- HANNIGAN, S. L. & REINITZ, M. T. (2001): A Demonstration and Comparison of Two Types of Interference-Based Memory Errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27 (4), 931-940.
- HASHTROUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1990): Aging and qualitative characteristics of Memories for perceived and imagined complex events. *Psychology and Aging*, 5, 119-126.
- HEKKANEN, S. T. & McEVOY, C. (2002): False Memories and Source-Monitoring Problems: Criterion Differences. *Applied Cognitive Psychology*, 16, 73-85.
- HENKEL, L. A. (2004): Erroneous Memories arising from Repeated Attempts to Remember. *Journal of Memory and Language*, 50, 26-46.
- HENKEL, L. A. & COFFMAN, K. J. (2004): Memory Distortions in Coerced False Confessions: A Source Monitoring Framework Analysis. *Applied Cognitive Psychology*, 18, 567-588.
- HENKEL, L. A., FRANKLIN, N. & JOHNSON, M. K. (2000): Cross-Modal Source Monitoring Confusions between Perceived and Imagined Events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26 (2), 321-335.
- HENKEL, L. A., JOHNSON, M. K. & De LEONARDIS, D. M. (1998): Aging and Source Monitoring : Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, 127, 251-268.
- HENRY, L. A. & GUDJONSSON, G. H. (1999): Eyewitness Memory and Suggestibility in Children with Mental Retardation. *American Journal on Mental Retardation*, 104 (6), 491-508.
- HENRY, L. A. & GUDJONSSON, G. H. (2003). Eyewitness Memory, Suggestibility and repeated Recall Sessions in Children with Mild and Moderate Intellectual Disabilities. *Law and Human Behavior*, 27 (5), 481-505.

- HENRY, L. A. & GUDJONSSON, G. H. (2004): The Effects of Memory Trace Strength on Eyewitness Recall in Children with and without Intellectual Disabilities. *Journal of Experimental Child Psychology*, 89, 53-71.
- HERSHKOWITZ, I., LAMB, M. E., STERNBERG, K. J. & ESPLIN, P. W., (1997): The Relationships Among Interviewer Utterance Type, CBCA Scores And The Richness Of Children's Responses. *Legal And Criminological Psychology*, 2, pp. 169-176.
- HERSHKOWITZ, I., ORBACH, Y., LAMB, M., STERNBERG, K. J. & HOROWITZ, D. (2001): The Effects of Mental Context Reinstatement on Children's Accounts of Sexual Abuse. *Applied Cognitive Psychology*, 15, 235-248.
- HICKS, J. L. & MARSH, R. L. (1999): Attempts to Reduce the Incidence of False Recall with Source Monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25 (5), 1195-1209.
- HICKS, J. L. & MARSH, R. L. (2001): False Recognition occurs more frequently during Source Identification than during Old-New Recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27 (2), 375-383.
- HICKS, J. L. & MARSH, R. L. & RITSCHER, L. (2002): The Role of Recollection and Partial Information in Source Monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28 (3), 503-508.
- HÖFER, E. & KÖHNKEN, G. (1998): Zeugen. In: Strauss, B.: *Zuschauer*. pp. 232-270. Göttingen: Hogrefe.
- HÖFER, E., KÖHNKEN, G., HAHNEWINKEL, R., BRUHN, C. (1993): *Diagnostik und Attribution von Glaubwürdigkeit*. Unveröffentlichter Schlussbericht für Grant 882 / 4-1 der Deutschen Forschungsgemeinschaft, Universität Kiel.
- HOFFMAN, H. G. (1997): Role of Memory Strength in Reality Monitoring Decisions: Evidence from Source Attribution Biases. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23 (2), 371-383.

- HOROWITZ, S. W., LAMB, M. E., ESPLIN, P. W., BOYCHUK, T. D., KRISPIN, O. & REITER-LAVERY, L. (1997): Reliability Of Criteria-Based Content Analysis Of Child Witness Statements. *Legal And Criminological Psychology*, 2, pp. 11-21
- HULME, C. & MacKENZIE, C. (1992): *Working Memory and Severe Learning Difficulties – Essays in Cognitive Psychology*. Hove: Lawrence Erlbaum Ass.
- HUNT, J. S. & BORGIDA, E. (2001): Is that what I said? Witnesses' Responses to Interviewer Modifications. *Law and Human Behavior*, 25 (6), 583-603.
- HYMAN, I. E., JR. & PENTLAND, J. (1996): The Role of Mental Imagery in the Creation of False Childhood Memories. *Journal of Memory and Language*, 35, 101-117.
- IHLEBAEK, C., LØVE, T., EILERTSEN, D. A. & MAGNUSSEN, S. (2003): Memory for a staged criminal event witnessed live and on video. *Memory*, 11 (3), 319-327.
- JOFFE, R. & YUILLE (1992): *Criteria-Based Content Analysis: An Experimental Investigation*. Paper presented at the NATO Advanced Study Institute: The Child Witness in Context – Cognitive, Social and Legal Perspectives, July 1992, Il Ciocco, Lucca, Italy.
- JOHNSON, M. K. (1988): Reality Monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, 117, 390-394.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988): Phenomenal Characteristics of Memories for perceived and imagined Autobiographical Events. *Journal of Experimental Psychology: General*, 117, 371-376.
- JOHNSON, M. K., HASHTROUDI, S. & LINDSAY, D. S. (1993): Source Monitoring. *Psychological Bulletin*, 114 (1), 3-28.
- JOHNSON, M. K. & MULTHAUP, K. S. (1992): Emotion and MEM. In: Christianson, S. A. (Ed.): *The Handbook of Emotion and Memory: Current research and theory*. pp. 33-66. Hillsdale, NJ: Erlbaum.
- JOHNSON, M. K., NOLDE, S. F. & DE LEONARDIS, D. M. (1996): Emotional Focus and Source Monitoring. *Journal of Memory and Language*, 35, 135-156.

- JOSLYN, S., LOFTUS, E., McNOUGHTON, A. & POWERS, J. (2001): Memory for Memory. *Memory & Cognition*, 29 (6), 789-797.
- JOYCE, T. A. (2003): An Audit of Investigations into Allegations of Abuse involving Adults with Intellectual Disability. *Journal of Intellectual Disability Research*, 47 (8), 606-616.
- KAIL, R. (1992): *Gedächtnisentwicklung bei Kindern*. Heidelberg, Berlin, New York: Spektrum Akad. Verlag.
- KAUSLER, D. H. (1994): *Learning and Memory in Normal Aging*. Academic Press: San Diego, CA.
- KEBBELL, M. R., HATTON, C., JOHNSON, S. D. & O'KELLY, C. M. E. (2001): People with Learning Disabilities as witnesses in court: What questions should lawyers ask? *British Journal of Learning Disabilities*, 29, 98-102.
- KEBBELL, M. R., HATTON, C. & JOHNSON, S. D. (2004): Witnesses with Intellectual Disabilities in court: What questions are asked and what influence do they have? *Legal and Criminological Psychology*, 9, 23-35.
- KEBBELL, M. R. & JOHNSON, S. D. (2000): Lawyer's Questioning: The Effect of Confusing Questions on Witness Confidence and Accuracy. *Law and Human Behavior*, 24 (6), 629-641.
- KELLEY, C. M. & SAHAKYAN, L. (2003): Memory, Monitoring and Control in the Attainment of Memory Accuracy. *Journal of Memory and Language*, 48, 704-721.
- KELLOWAY, E. K., STINTSON, V. & MacLEAN, C. (2004): Eyewitness Testimony in Occupational Accident Investigations: Towards a Research Agenda. *Law and Human Behavior*, 28 (1), 115-132.
- KELLY, A., CARROLL, M. & MAZZONI, G. (2002): Metamemory and Reality Monitoring. *Applied Cognitive Psychology*, 16, 407-428.
- KLJAJIC, T. (2003). *Sexuelle Gewalt an Menschen mit geistiger Behinderung und die Relevanz für die Soziale Arbeit*. Unveröffentlichte Diplomarbeit. Universität Eichstätt.

- KÖHNKEN, G. (1999A): Glaubwürdigkeit. In: Lempp, R., Schütze, G. & Köhnken, G. (Hrsg.): *Forensische Psychiatrie und Psychologie des Kindes- und Jugendalters*. pp. 318-341. Darmstadt: Steinkopff.
- KÖHNKEN, G. (1999B): Suggestion und Suggestibilität. In: Lempp, R., Schütze, G. & Köhnken, G. (Hrsg.): *Forensische Psychiatrie und Psychologie des Kindes- und Jugendalters*. pp. 342-353. Darmstadt: Steinkopff.
- KÖHNKEN, G. (2001): Methodik der Glaubwürdigkeitsbegutachtung. In: Fegert, J. (Hrsg.): *Begutachtung sexuell missbrauchter Kinder*. pp. 29-51. Neuwied; Kriftel: Luchterhand.
- KÖHNKEN, G., SCHIMOSSEK, E., ASCHERMANN, E. & HÖFER, E. (1995): The Cognitive Interview and the Assessment of the Credibility of Adults' Statements. *Journal of Applied Psychology*, 80 (6), 671-684.
- KORIAT, A., GOLDSMITH, M., SCHNEIDER, W. & NAKASH-DURA, M. (2001): The Credibility of Children's Testimony: Can Children Control their Accuracy of their Memory Reports? *Journal of Experimental Child Psychology*, 79, 405-437.
- KRAHÉ, B. & KUNDROTAS, S. (1992): Glaubwürdigkeitsbeurteilung bei Vergewaltigungsanzeigen: Ein aussageanalytisches Feldexperiment. *Zeitschrift für experimentelle und angewandte Psychologie*, 39, 598-620.
- KRINSKY-McHALE, S. J., DEVENNY, D. A., KITTLER, P. & SILVERMAN, W. (2003): Implicit Memory in Aging Adults with Mental Retardation with and without Down Syndrome. *American Journal on Mental Retardation*, 108 (4), 219-233.
- KRUPSKI, A. (1986): Attention Problems in Youngsters with Learning Handicaps. In: Torgesen, J. K. & Wong, B. Y. L. (Eds.): *Psychological and Educational Perspectives on Learning Disabilities*. pp. 161-192. Orlando: Academic Press.
- KVAM, M. H. (2000): Is Sexual Abuse of Children with Disabilities disclosed? *Child Abuse and Neglect*, 24, 1073 – 1084.
- LA ROOY, D., PIPE, M.-E. & MURRAY, J. E. (2005): Reminiscence and Hypermnnesia in Children's Eyewitness Memory. *Journal of Experimental Child Psychology*, 90, 235-254.

- LAMB, M. E. & FAUCHIER, A. (2001): The Effects of Question Type on Self-Contradictions by Children in the Course of Forensic Interviews. *Applied Cognitive Psychology*, 15, 483-491.
- LAMB, M. E., STERNBERG, K. J., ESPLIN, P. W., HERSHKOWITZ, I., ORBACH, Y. & HOVAV, M. (1997): Assessing The Credibility Of Children's Allegations Of Sexual Abuse: A Survey Of Recent Research. *Learning And Individual Differences*, 9 (2), 175-194.
- LAMB, M., STERNBERG, K., ESPLIN, P., HERSHKOWITZ, I., ORBACH, Y. & HOVAV, M. (1997): Criterion-Based Content Analysis: A Field Validation Study. *Child Abuse And Neglect*, 21 (3), 255-264.
- LAMERS-WINKELMAN, F. & BUFFING, F. (1996): Children's Testimony In The Netherlands: A Study Of Statement Validity Analysis. *Criminal Justice And Behavior*, 23 (2), 304-321.
- LAMPINEN, J. M., ODEGARD, T. N. & BULLINGTON, J. L. (2003): Qualities of Memories for Performed and Imagined Actions. *Applied Cognitive Psychology*, 17, 881-893.
- LANDRY, K. L. & BRIGHAM, J. C. (1992): The Effect of Training In Criteria-Based Content Analysis On The Ability To Detect Deception In Adults. *Law and Human Behavior*, 10 (6), 663-676.
- LEHMAN, E. B. et al. (2001): Item-Cued directed Forgetting of related words and pictures in Children and Adults: Selective Rehearsal versus Cognitive Inhibition. *The Journal of General Psychology*, 128 (1), 81-97.
- LIBBY, L. K. (2003): Imagery Perspective and Source Monitoring in Imagination Inflation. *Memory & Cognition*, 31 (7), 1072-1081.
- LICHT, B. G. & KISTNER, J. A. (1986): Motivational Problems of Learning-Disabled Children: Individual Differences and their Implications for Treatment. In: Torgesen, J. K. & Wong, B. Y. L. (Eds.): *Psychological and Educational Perspectives on Learning Disabilities*. pp. 225-256. Orlando: Academic Press.

- LINDSAY, D. S. (1990): Misleading questions can impair Eyewitness' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1077-1083.
- LINDSAY, D. S. (1994): Memory Source Monitoring and Eyewitness Testimony. In: Ross, D. F., Read, J. D. & Toglia, M. P. (Eds.): *Adult Eyewitness Testimony – Current Trends and Developments*. pp. 27-55. Cambridge: University Press.
- LINDSAY, D. S. (2002): Children's Source Monitoring. In: Westcott, H. L., Davies, G. M. & Bull, R. (Eds.): *Children's Testimony. A Handbook of Psychological Research and Forensic Practice*. pp. 83-98. Chichester: Wiley.
- LINDSAY, D. S., ALLEN, B. P., CHAN, J. C.K. & DAHL, L. (2004): Eyewitness Suggestibility and Source Similarity: Intrusions of Details from one Event into Memory Reports of another Event. *Journal of Memory and Language*, 50, 96-111.
- LINDSAY, D. S., & JOHNSON, M. K. (1989): The Eyewitness Suggestibility Effect and Memory for Source. *Memory & Cognition*, 17 (3), 349–358.
- LINDSAY, D. S., JOHNSON, M. K. & KWON, P. (1991): Developmental Changes in Memory Source Monitoring. *Journal of Experimental Child Psychology*, 52, 297-318.
- LINDSAY, W. R. & BEAIL, N. (2004): Risk Assessment: Actuarial Prediction and Clinical Judgement of Offending Incidents and Behaviour for Intellectual Disability Services. *Journal of Applied Research in Intellectual Disabilities*, 17, 229-234.
- LIPSCOMB, T. J., McALLISTER, H. A. & BREGMAN, N. (1985): Bias in Eyewitness Accounts: The Effects of Question Format, Delay Interval and Stimulus Presentation. *The Journal of Psychology*, 119 (3), 207-212.
- LOFTUS, E. F. (1979): The Malleability of Memory. *American Scientist*, 67, 312-320.
- LORSBACH, T. C. & EWING, R. H. (1995): Source Monitoring in Children with Learning Disabilities. *International Journal of Disability, Development and Education*, 42, 241-257.

- LORSBACH, T. C. (2000): Source Monitoring as a Framework for Conceptualizing the Nature of Memory Difficulties in Children with Learning Disabilities. In: Roberts, K. P. & Blades, M. (Eds.): *Children's Source Monitoring*. pp. 115-146. Mahwah: Lawrence Erlbaum Associates.
- LUCKASSON, R. (1992). People with Mental Retardation as Victims of Crime. In: Conley, R. W., Luckasson, R. & Bouthilet, G. N. (Eds.): *The Criminal Justice System and Mental Retardation – Defendants and Victims*, pp. 209-220. Baltimore: Paul H. Brooks.
- MÄHLER, C. & HASSELHORN, M. (1990): Gedächtnisdefizite bei lernbehinderten Kindern: Entwicklungsverzögerung oder Strukturdifferenz? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22 (4), 354-366.
- MAGLIANO, J. P., MILLER, J. & ZWAAN, R. A. (2001): Indexing Space and Time in Film Understanding. *Applied Cognitive Psychology*, 15, 533-545.
- MAPES, B. E. (1995): *Child Eyewitness Testimony in Sexual Abuse Investigations*. Brandon, Vermont: Clinical Psychology Publishing Company.
- MARKHAM, R. & HYNES, L. (1993): The effect of vividness of imagery on Reality Monitoring. *Journal of Mental Imagery*, 17 (3-4), 159-170.
- MARSH, R. L. & HICKS, K. L. (1998): Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1137-1151.
- MARSH, R. L. & HICKS, J. L. (2002): Comparisons of Target Output Monitoring and Source Input Monitoring. *Applied Cognitive Psychology*, 16, 845-862.
- MARSH, R. L., HICKS, J. L. & DAVIS, T. T. (2002): Source Monitoring does not alleviate (and may exacerbate) the Occurrence of Memory Conjunction Errors. *Journal of Memory and Language*, 47, 315-326.
- MARSH, E. J., MEADE, M. L. & ROEDIGER, H. L. (2003): Learning Facts from Fiction. *Journal of Memory and Language*, 49, 519-536.

- MARSH, E. J. & TVERSKY, B. (2004): Spinning the Stories of our Lives. *Applied Cognitive Psychology*, 18, 491-503.
- MARSH, E. J., TVERSKY, B. & HUTSON, M. (2005): How Eyewitnesses Talk about Events: Implications for Memory. *Applied Cognitive Psychology*, ??, in press.
- MATHER, M., SHAFIR, E. & JOHNSON, M. K. (2000): Misremembrance of Options Past: Source Monitoring and Choice. *Psychological Science*, 11 (2), 132-138.
- MCCLOSKEY, M., & ZARAGOZA, M. (1985): Misleading postevent information and Memory for Events: Arguments and evidence against memory impairment hypothesis. *Journal of Experimental Psychology: General*, 114, 1-16.
- McNAMARA, J. K. & WONG, B. (2003): Memory for Everyday Information in Students with Learning Disabilities. *Journal of Learning Disabilities*, 36 (5), 394-406.
- MEISER, T. & BRÖDER, A. (2002): Memory for Multidimensional Source Information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28 (1), 116-137.
- MEMON, A., HOPE, L., BARTLETT, J. & BULL, R. (2002): Eyewitness Recognition Errors: The Effect of mugshot viewing and choosing in young and old adults. *Memory & Cognition*, 30 (8), 1219-1227.
- MEYER-GOßNER, L. (2001): *Strafprozessordnung. GVG, Nebengesetze und ergänzende Bestimmungen*. 45. Auflage. München: Beck.
- MILNE, R., CLARE, I. & Bull, R. (1996, 1999): Using the Cognitive Interview with Adults with Mild Learning Disability. *Psychology, Crime, and Law*, 5, 81 – 101.
- MILNE, R. (1999): Interviewing Children with Learning Disabilities. In: Memon, A. & Bull, R. (Eds.): *Handbook of the Psychology of Interviewing*, pp. 165-180. Chichester: Wiley.
- MILNE, R. & BULL, R. (2001): Interviewing Witnesses with Learning Disabilities for Legal Purposes. *British Journal of Learning Disabilities*, 29, 93-97.
- MILNE, R. & BULL, R. (2003). *Psychologie der Vernehmung. Die Befragung von Tatverdächtigen, Zeugen und Opfern*. Bern: Verlag Hans Huber.

- MILNE, R., CLARE, I. & BULL, R. (1996, 1999): Using the Cognitive Interview with Adults with Mild Learning Disability. *Psychology, Crime and Law*, 5, 81 – 101.
- MILNE, R., CLARE, I. & BULL, R. (2002): Interrogative Suggestibility among Witnesses with Mild Intellectual Disabilities: The Use of an Adaption of the GSS. *Journal of Applied Research in Intellectual Disabilities*, 15, 8-17.
- MITCHELL, K. J. & JOHNSON, M. K. (2000): Source Monitoring: Attributing Mental Experiences. In: Tulving, E. & Craik, F. I. M. (Eds.): *The Oxford Handbook of Memory*, pp. 179-191. Oxford: University Press.
- MITCHELL, K., JOHNSON, M., & MATHER, M. (2003): Source Monitoring and Suggestibility to Misinformation: Adult Age-Related Differences. *Applied Cognitive Psychology*, 17, 107-119.
- MITCHELL, K. J. & ZARAGOZA, M. (2001): Contextual Overlap and Eyewitness Suggestibility. *Memory & Cognition*, 29 (4), 616-626.
- MÖLLER, H. J., LAUX, G. & DEISTER, A. (2001): *Psychiatrie und Psychotherapie*. Stuttgart: Thieme.
- MÜLLER-LUCKMANN, E. (1980). Beurteilung der Glaubhaftigkeit von Zeugenaussagen. In: Schneider, H. J. (Hrsg.), *Psychologie des 20. Jahrhunderts, Band 14: Auswirkungen auf die Kriminologie*, 791-825. Zürich: Kindler.
- MUSSWEILER, T. & NEUMANN, R. (2000): Sources of Mental Contamination: Comparing the Effects of Self-Generated versus Externally Provided Primes. *Journal of Experimental Social Psychology*, 36, 194-206.
- NARBY, D. J., CUTLER, B. L. & PENROD, S. D. (1996): The Effects of Witness, Target and Situational Factors on Eyewitness Identifications. In: Sporer, S. L., Malpass, R. S. & Köhnken, G. (Eds.): *Psychological Issues in Eyewitness Identification*. pp. 23-52. Mahwah: Lawrence Erlbaum Associates.
- NAVEH-BENJAMIN, M., HUSSAIN, Z., GUEZ, J. & BAR-ON, M. (2003): Adult Age Differences in Episodic Memory: Further Support for an Associative-Deficit Hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29 (5), 826-837.

- NIEDZWIENSKA, A. (2003): Distortion of Autobiographical Memories. *Applied Cognitive Psychology*, 17, 81-91.
- NEUHÄUSER, G. & STEINHAUSEN, H. C. (1999). *Geistige Behinderung. Grundlagen, Klinische Syndrome, Behandlung und Rehabilitation*. Zweite überarbeitete Auflage. Stuttgart: Kohlhammer.
- NOAK, C. & SCHMID, H. J. (1994). *Sexuelle Gewalt gegen Menschen mit geistiger Behinderung. Eine verleugnerte Realität*. Stuttgart: Klett-Verlag.
- OPP, G. (1992): A German Perspective on Learning Disabilities. *Journal of Learning Disabilities*, 25 (6), 351-360.
- ORBACH, Y. & LAMB, M. E. (1999): Assessing The Accuracy Of A Child's Account Of Sexual Abuse: A Case Study. *Child Abuse And Neglect*, 23 (1), 91-98.
- OST, J., VRIJ, A., COSTALL, A. & BULL, R. (2002): Crashing Memories and Reality Monitoring: Distinguishing between Perceptions, Imaginations and 'False Memories'. *Applied Cognitive Psychology*, 16, 125-134.
- PARKER, A. P. & BROWN, J. (2000): Detection Of Deception: Statement Validity Analysis As A Means Of Determining Truthfulness Or Falsity Of Rape Allegations. *Applied Developmental Science*, 5, 237-259.
- PELED, M. IAROCCHI, G. & CONNOLLY, D. A. (2004): Eyewitness Testimony and Perceived Credibility of Youth with Mild Intellectual Disability. *Journal of Intellectual Disability Research*, 48 (7), 699-703.
- PERLMAN, N. B., ERICSON, K. I., ESSES, V. M. & ISAACS, B. J. (1994): The Developmentally Handicapped Witness. Competency as a Function of Question Format. *Law and Human Behavior*, 18 (2), 171-187.
- PETERSON, C. & WHALEN, N. (2001): Five Years Later: Children's Memory for Medical Emergencies. *Applied Cognitive Psychology*, 15, 7-24.

- PIPE, M.-E., LAMB, M. E., ORBACH, Y. & ESPLIN, P. W. (2004): Recent Research on Children's Testimony about Experienced and Witnessed Events. *Developmental Review*, 24, 440-468.
- POLCZYK, R., WESOŁOWSKA, B., GABARCZYK, A., MINAKOWSKA, I., SUPSKA, M. & BOMBA, E. (2004): Age Difference in Interrogative Suggestibility: A Comparison between Young and Older Adults. *Applied Cognitive Psychology*, 18, 1097-1107.
- POLLOWAY, E. A., PATTON, J. R., SMITH, T. E. C. & BUCK, G. H. (1997): Mental Retardation and Learning Disabilities: Conceptual and Applied Issues. *Journal of Learning Disabilities*, 30 (3), 297-308.
- POOLE, D. A. & WHITE, L. T. (1991): *Effects of question repetition and retention interval on the eyewitness testimony of children and adults*. *Developmental Psychology*, 27, 975-986.
- POOLE, D. A. & WHITE, L. T. (1993): Two years later: Effects of Question Repetition and Retention Interval on the Eyewitness Testimony of Children and Adults. *Developmental Psychology*, 29 (5), 844-853.
- POOLE, D. A. & WHITE, L. T. (1995): Tell me again and again: The stability and change in the repeated testimony and adults. In: Zaragoza, M. S., Graham, J. R., Hall, G., Hirschman, R. & Ben-Porath, Y. S. (Eds.): *Memory and Testimony in the Child Witness*. pp. 24-43. Thousand Oaks, CA: Sage.
- POWELL, M. B. & THOMSON, D. M. (2003): Improving Children's Recall of an Occurrence of a Repeated Event: Is it a matter of helping them to generate options? *Law and Human Behavior*, 27 (4), 365-384.
- PRESSLEY, M. (1991): Can Learning-Disabled Children become good Information Processors? How can we find out? In: Feagans, L. V., Short, E. J. & Meltzer, L. (Eds.): *Subtypes of Learning Disabilities: Theoretical perspectives and research*. pp. 137-162. Hillsdale: Erlbaum.
- QUAS, J. A., SCHAAF, J. M., ALEXANDER, K. W. & GOODMAN, G. S. (2000): Do you really remember it happening or do you only remember being asked about it happening? Children's Source Monitoring in Forensic Contexts. In: Roberts, K. P. & Blades, M.

- (Eds.): *Children's Source Monitoring*. pp. 197-226. Mahwah: Lawrence Erlbaum Associates.
- RANTZEN, A. & MARKHAM, R. (1992): The Reversed Eyewitness Testimony Design: More Evidence for Source Monitoring. *The Journal of General Psychology*, 119 (1), 37-43.
- RAPP, A. F. (1999): Wie "gut" erinnern Kinder im Vorschulalter Ereignisse in Abhängigkeit von unterschiedlichen Interviewtechniken? *Monatszeitschrift für Kriminologie und Strafrechtsreform*, 82 (3), 160-177.
- RAT für KRIMINALITÄTSVERHÜETUNG SCHLESWIG-HOLSTEIN (Hrsg.) - (2003). *Konzept zur Kriminalitätsverhütung. Gewalt gegen Menschen mit Behinderung: Abschlussbericht der Arbeitsgruppe 14*. Kiel.
- RATNER, H. H., FOLEY, M. A. & GIMPERT, N. (2000): Person Perspectives on Children's Memory and Learning: What do Source-Monitoring Failures Reveal? In: Roberts, K. P. & Blades, M. (Eds.): *Children's Source Monitoring*. pp. 85-114. Mahwah: Lawrence Erlbaum Associates.
- REYNA, V. F., HOLLIDAY, R. & MARCHE, T. (2002): Explaining the Development of False Memories. *Developmental Review*, 22, 436-489.
- ROEBERS, C. & SCHNEIDER, W. (2001A): Individual Differences in Children's Eyewitness Recall: The Influence of Intelligence and Shyness. *Applied Developmental Science*, 5 (1), 9-20.
- ROEBERS, C. & SCHNEIDER, W. (2001B): Memory for an Observed Event in the Presence of Prior Misinformation: Developmental Patterns of Free Recall and Identification Accuracy. *British Journal of Developmental Psychology*, 19, 507-524.
- ROBERTS, K. P. (2000): Introduction: Children's Source Monitoring. In: Roberts, K. P. & Blades, M. (Eds.): *Children's Source Monitoring*. pp. 1-11. Mahwah: Lawrence Erlbaum Associates.
- ROBERTS, K. P. (2002): Children's Ability to distinguish between Memories from Multiple Sources: Implications for the Quality and Accuracy of Eyewitness Statements. *Developmental Review*, 22, 403-435.

- ROBERTS, P. (2002): Vulnerability to False Memory: The Effects of Stress, Imagery, Trait Anxiety and Depression. *Current Psychology: Developmental – Learning – Personality – Social*, 21 (3), 240-252.
- ROBERTS, K. P. & BLADES, M. (1998): The Effects of Interacting with Events on Children's Memory and Source Monitoring. *Applied Cognitive Psychology*, 12, 489-503.
- ROTH, M. (1997): *Beiträge zu Gedächtnisleistungen bei Kindern mit allgemeinen spezifischen Lernstörungen am Beispiel der Lernbehinderung und der umschriebenen Lese-Rechtschreibschwäche*. Inaugural-Dissertation. Universität Würzburg.
- RUBIN, D. C. & BERNTSEN, D. (2003): Life Scripts help to maintain Autobiographical Memories of highly positive, but not highly negative events. *Memory & Cognition*, 31 (1), 1-14.
- RUBY, C. L. & BRIGHAM, J. C. (1997): The Usefulness Of The Criteria-Based Content Analysis Technique In Distinguishing Between Truthful And Fabricated Allegations: A Critical Review. *Psychology, Public Policy and Law*, 3 (4), 705-737.
- RUBY, C. L. & BRIGHAM, J. C. (1998): Can Criteria-Based Content Analysis Distinguish Between True And False Statements Of African-American Speakers? *Law and Human Behavior*, 22 (4), 369-388.
- SALOVEY, P., SIEBER, W. J., JOBE, J. B. & WILLIS, G. B. (1994): The Recall of Physical Pain. In: Schwarz, N. & Sudman, S. (Eds.): *Autobiographical Memory and the Validity of Retrospective Reports*. pp. 89-106. New York: Springer.
- SCHLEIDER, K. (1993): *Komponenten der Gedächtnisentwicklung – Beiträge zum Entwicklungsverlauf und dem Zusammenspiel von Kapazität, Strategiegebrauch und Strategiewissen unter Berücksichtigung von Reproduktions- und Schulleistung*. Frankfurt am Main: Peter Lang.
- SCHNEIDER, W. & PRESSLEY, M. (Eds.) - (1997): *Memory Development between two and twenty*. Mahwah: Erlbaum.
- SCHOPF, P. (1998). Lernschwäche, Lernbeeinträchtigung, Lernbehinderung: Aspekte zur Klärung. Eine Arbeitshilfe für eine differenzierte Betrachtungsweise und ein besseres

- Verständnis, aber auch eine Anregung für die Weiterentwicklung. In *Informationen für die Beratungs- und Vermittlungsdienste der Bundesanstalt für Arbeit (ibv)*, 6/98, (S. 341-449). Nürnberg.
- SCHÜLER, (1997): *Projekt: Untersuchung von Sexualstraftaten*. Dokumentation Nr. 17 des Ausbildungszentrums für Verwaltung, Verwaltungsfachhochschule, Fachbereich Polizei.
- SCHUMAKER, J. B., DESHLER, D. D. & ELLIS, E. S. (1986): Intervention Issues related to the Education of LD Adolescents. In: Torgesen, J. K. & Wong, B. Y. L. (Eds.): *Psychological and Educational Perspectives on Learning Disabilities*. pp. 329-366. Orlando: Academic Press.
- SCHWARZ, S., ROEBERS, C. & SCHNEIDER, W. (2004): Entwicklungsveränderungen in Konformität und in kognitiven Folgen sozialer Beeinflussung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36 (4), 173-181.
- SCHWEIZER, K. (1996): Level of Encoding, Preattentive Processing and Working-Memory Capacity as Sources of Cognitive Ability. *Personality and Individual Differences*, 21 (5), 759-766.
- SEAMON, J. G. et al. (2002): Are False Memories more difficult to forget than accurate Memories? The effect of Retention Interval on Recall and Recognition. *Memory & Cognition*, 30 (7), 1054-1064.
- SENN, C. Y. (1993). *Gegen jedes Recht: Sexueller Missbrauch und geistige Behinderung*. Berlin: Donna Vita.
- SHAPIRO, L. R. & PURDY, T. L. (2005): Suggestibility and Source Monitoring Errors: Blame the Interview Style, Interviewer Consistency and the Child's Personality. *Applied Cognitive Psychology*, ??, in press.
- SHAW, J. S., GARVEN, S. & WOOD, J. M. (1997): Co-Witness Information can have Immediate Effects on Eyewitness Memory Reports. *Law and Human Behavior*, 21 (5), 503-523.

- SHAW, J. S. & ZERR, T. K. (2003): Extra Effort during Memory Retrieval may be associated with Increases in Eyewitness Confidence. *Law and Human Behavior*, 27 (3), 315-329.
- SHERMAN, J. W. & BESSENOFF, G. R. (1999): Stereotypes as Source-Monitoring Cues: On the Interaction between Episodic and Semantic Memory. *Psychological Science*, 10 (2), 106-110.
- SMITH, S. M., TINDELL, D. R., PIERCE, B. H., GILLILAND, T. R. & GERKENS, D. R. (2001): The Use of Source Memory to Identify One's Own Episodic Confusion Errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27 (2), 362-374.
- SOBSEY, D. (1994): Sexual Abuse of Individuals with Intellectual Disability. In: Craft, A. (Ed.): *Practice Issues in Sexuality and Learning Disabilities*. pp. 93-115. London: Routhledge.
- SPANIOL, J. & BAYEN, U. J. (2002): When is Schematic Knowledge Used in Source Monitoring? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28 (4), 631-651.
- SPENCER, W. D. & RAZ, N. (1995): Differential effects of aging on Memory for content and context: A metaanalysis. *Psychology and Aging*, 10, 527-539.
- SPOKER, S. L. (1996): Psychological Aspects of Person Descriptions. In: Sporer, S. L., Malpass, R. S. & Köhnken, G. (Eds.): *Psychological Issues in Eyewitness Identification*. pp. 53-86. Mahwah: Lawrence Erlbaum Associates.
- SPOKER, S. L. (1997A): Realitätsüberwachungskriterien und forensische Glaubwürdigkeitskriterien im Vergleich: Validitätsüberprüfungen anhand selbsterlebter und erfundener Geschichten. In: Greuel, L., Fabian, T. & Stadler, M. (Hrsg.): *Psychologie der Zeugenaussage*. pp. 71-86. Weinheim: Psychologie Verlags Union.
- SPOKER, S. L. (1997B): The Less Travelled Road To Truth: Verbal Cues In Deception Detection In Accounts Of Fabricated And Self-Experienced Events. *Applied Cognitive Psychology*, Vol. 11, pp. 373-397.

- SPODER, S. L. & BURSCHE, S. E. (1997): Kinder vor Gericht: Soziale und kognitive Voraussetzungen der Aussagen von Kindern. In: *Psychologische Rundschau*, 48, pp. 141-162.
- STELLER, M. (1988): Die vierte Phase der Aussagepsychologie. *Forensia* 9, pp. 23-28.
- STELLER, M. (1998): Aussagepsychologie vor Gericht. Methodik und Probleme von Glaubwürdigkeitsgutachten mit Hinweisen auf die Wormser Missbrauchsprozesse. *Recht und Psychiatrie*, 1, xx-xx.
- STELLER, M. & VOLBERT, R. (1997): Glaubwürdigkeitsbegutachtung. In: Steller, M. & Volbert, R. (Hrsg.): *Psychologie im Strafverfahren: Ein Handbuch*. pp. 12-39. Bern: Huber.
- STELLER, M. & VOLBERT, R. (1999): *Wissenschaftliches Gutachten. Forensisch-aussagepsychologische Begutachtung (Glaubwürdigkeitsbegutachtung)*. Berlin: Institut für Forensische Psychiatrie der Freien Universität Berlin.
- STELLER, M., WELLERSHAUS, P. & WOLF, T. (1992): Realkennzeichen in Kinderaussagen: Empirische Grundlagen der Kriterienorientierten Aussageanalyse. In: *Zeitschrift für experimentelle und angewandte Psychologie*, 39 (1), pp. 151-171.
- STERN, L. B. & DUNNING, D. (1994): Distinguishing accurate from inaccurate Eyewitness Identifications: A Reality Monitoring Approach. In: Ross, D. F., Read, J. D. & Toglia, M. P. (Eds.): *Adult Eyewitness Testimony – Current Trends and Developments*. pp. 273-299. Cambridge: University Press.
- STEWART, M. S. (1991): Understanding Children's Memories of Medical Procedures: "He didn't touch me and it didn't hurt!". In: Nelson, C. A. (Ed.): *Memory and Affect in Development*. pp. 171-227. Hillsdale: Erlbaum.
- STOBBS, G. & KEBBEL, M. R. (2003): Juror's Perception of Witnesses with Intellectual Disabilities and the Influence of Expert Evidence. *Journal of Applied Research in Intellectual Disabilities*, 16, 107-114.
- STROMSNESS, M. M. (1993): Sexually Abused Women with Mental Retardation: Hidden Victims, Absent Resources. *Women and Therapy*, 14 (3-4), 139-152.

- STRÖMWALL, L. A., BENGTSSON, L., LEANDER, L. & GRANHAG, P. A. (2004): Assessing Children's Statements: The Impact of a Repeated Experience on CBCA and RM Ratings. *Applied Cognitive Psychology*, 18, 653-668.
- SUENGAS, A. G. & JOHNSON, M. K. (1988): Qualitative effects of rehearsal in Memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, 117, 377-389.
- SULLIVAN, P. M. & KNUTSON, J. F. (2000): Maltreatment and Disabilities: a population-based epidemiological study. *Child Abuse and Neglect*, 24, 1257 – 1273.
- SUSSMAN, A. L. (2001): Reality Monitoring of Performed and Imagined Interactive Events: Developmental and Contextual Effects. *Journal of Experimental Child Psychology*, 79, 115-138.
- SUTHERLAND, R. & HAYNE, H. (2001): The Effect of Postevent Information on Adults' Eyewitness Reports. *Applied Cognitive Psychology*, 15, 249-263.
- THIERRY, K. L. & SPENCE, M. J. (2004): A Real-Life Event Enhances the Accuracy of preschoolers' Recall. *Applied Cognitive Psychology*, 18, 297-309.
- TUCKEY, M. R. & BREWER, N. (2003): The Influence of Schemas, Stimulus Ambiguity and Interview Schedule on Eyewitness Memory over time. *Journal of Experimental Psychology: Applied*, 9 (2), 101-118.
- TULVING, E. (2000): Concepts of Memory. In: Tulving, E. & Craik, F. I. M. (Eds.): *The Oxford Handbook of Memory*, pp. 33-43. Oxford: University Press.
- TYE, M. C., AMATO, S. L., HONTS, C. R., DEVITT, M. K. & PETERS, D. (1999): The Willingness Of Children To Lie And The Assessment Of Credibility In An Ecologically Relevant Laboratory Setting, *Applied Developmental Science*, 3 (2), 92-109.
- VALENTI-HEIN, D. C. & SCHWARTZ, L. D. (1993): Witness Competency in People with Mental Retardation: Implications for Prosecution of Sexual Abuse. *Sexuality and Disability*, 11 (4), 287-294.

- VERDUGO, M. A. & BERMEJO, B. G. (1997): The Mentally Retarded Person as a Victim of Maltreatment. *Aggression and Violent Behavior*, 2 (2), 143-165.
- VICARI, S. & CARLESIMO, G. A. (2002): Children with Intellectual Disabilities. In: Baddeley, A. D., Kopelman, M. D. & Wilson, B. A. (Eds.): *Handbook of Memory Disorders*. New York: John Wiley & Sons Inc.
- VOLBERT, R. (1997): Suggestibilität kindlicher Zeugen. In: Steller, M. & Volbert, R. (Hrsg.): *Psychologie im Strafverfahren: Ein Handbuch*. pp. 40-62. Bern: Huber.
- VOLBERT, R. (2000). Standards der psychologischen Glaubhaftigkeitsdiagnostik. In: Kröber, H.-L. & Steller, M. (Hrsg.): *Psychologische Begutachtung im Strafverfahren. Indikationen, Methoden und Qualitätsstandards*, S. 113-147. Darmstadt: Steinkopff.
- WAFSTAFF, G. F. et al. (2003): Can Laboratory Findings on Eyewitness Testimony be generalized to the real world? An archival analysis of the Influence of Violence, Weapon Presence and Age on Eyewitness Accuracy. *The Journal of Psychology*, 137 (1), 17-28.
- WALLER, H. (1997): *Sozialmedizin*. Stuttgart: Kohlhammer.
- WALTER, J. (Hrsg.) - (1992): *Sexualität und geistige Behinderung*. Heidelberg: Schindele Verlag.
- WARD, G. & CARROLL, M. (1997): Reality Monitoring for Sexual Abuse Memories. *Applied Cognitive Psychology*, 11, 293-304.
- WATERMAN, A. H, BLADES, M. & SPENCER, C. (2001): Interviewing Children and Adults: The Effect of Question Format on the Tendency to Speculate. *Applied Cognitive Psychology*, 15, 521-531.
- WEIß, H.: *Lernbehinderung. Das Online-Familienhandbuch*.
(http://www.familienhandbuch.de/cmain/f_Aktuelles/a_Behinderung/s_586.html)
- WELCH-ROSS, M. K. (1995): An Integrative Model of the Development of Autobiographical Memory. *Developmental Review*, 15, 338-365.

- WELCH-ROSS, M. K. (2000): A Mental-State Reasoning Model of Suggestibility and Memory Source Monitoring. In: Roberts, K. P. & Blades, M. (Eds.): *Children's Source Monitoring*. pp. 227-256. Mahwah: Lawrence Erlbaum Associates.
- WELLS, G. L. & OLSON, E. A. (2003): Eyewitness Testimony. *Annual Review of Psychology*, 54, 277-295.
- WENDELER, J. (1993). *Geistige Behinderung. Pädagogische und psychologische Aufgaben*. Weinheim: Beltz Verlag.
- WINOGRAD, E., PELUSO, J. P. & GLOVER, T. A. (1998): Individual Differences in Susceptibility to Memory Illusions. *Applied Cognitive Psychology*, 12, 5-27.
- WIPPICH, W. (1984): *Lehrbuch der angewandten Gedächtnispsychologie - Bd.1*. Stuttgart: Kohlhammer.
- WIPPICH, W. (1985): *Lehrbuch der angewandten Gedächtnispsychologie - Bd.2*. Stuttgart: Kohlhammer.
- WOLF, P. & STELLER, M. (1997): Realkennzeichen in Aussagen von Frauen: Zur Validierung der kriterienorientierten Aussageanalyse für Zeugenaussagen von Vergewaltigungsopfern. In: Greuel, L, Fabian, T., Stadler, M. (Hrsg.): *Psychologie der Zeugen-aussage. Ergebnisse der rechtspsychologischen Forschung*. pp. 121-130. Weinheim: Psychologie Verlags Union.
- WOODY-DORNING, J. & MILLER, P. H. (2001): Children's Individual Differences in Capacity: Effects on Strategy Production and Utilization. *British Journal of Developmental Psychology*, 19, 543-557.
- WORDEN, P. E. (1983): Memory Strategy Instruction with the Learning Disabled. In: Pressley, M. & Levin, J. R. (Eds.): *Cognitive Strategy Research – Psychological Foundations*. pp. 129-151. New York: Springer.
- WRIGHT, D. B., LOFTUS, E. & HALL, M. (2001): Now you see it; now you don't: Inhibiting Recall and Recognition of Scenes. *Applied Cognitive Psychology*, 15, 471-482.

- WRIGHT, D. B. & STROUD, J. N. (2002): Age Differences in Line-Up Identification Accuracy: People are better with their own Age. *Law and Human Behavior*, 26 (6), 641-654.
- YUILLE, J. C. & TOLLESTRUP, P. A. (1992). A Model of the diverse effects of Emotion on Eyewitness Memory. In: Christianson, S. A. (Ed.): *The Handbook of Emotion and Memory*. Hillsdale, NJ: Erlbaum.
- ZAPARNIUK, J., YUILLE, J. C., TAYLOR, S. (1995): Assessing The Credibility Of True And False Statements. *International Journal of Law and Psychiatry*, 18 (3), 343-352.
- ZARAGOZA, M. S. & LANE, S. M. (1994): Source Misattributions and the Suggestibility of Eyewitness Memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20 (4), 934-945.
- ZELFEL, R. C. (2001). *Der Umgang mit dem Begriff „lernbehindert“ im Rahmen der beruflichen Rehabilitation, des ICIDH-2 der WHO und des SGB IX*. Referat am 23. März 2001 auf der Fachtagung „teil-haben.de – Berufliche Qualifizierung lernbehinderter Jugendlicher und ihre Eingliederung in die Berufs- und Arbeitswelt“. Berlin. Resource: www.teil-haben.de/html/download.html (Stand: 11. September 2004).
- ZEMP, A. & PIRCHER, E. (1996). *Weil das alles weh tut mit Gewalt. Sexuelle Ausbeutung von Mädchen und Frauen mit Behinderung*. Wien: Bundesministerium für Frauenangelegenheiten.
- ZIGLER, E. & BALLA, D. (1982) (Eds.): *Mental Retardation. The Developmental-Difference Controversy*. Hillsdale, New Jersey: Erlbaum.

ANHANGAnhang A: Instruktionen zur Fragestellung der Erinnerungsleistung und dem source- bzw. reality monitoring

- *Allgemeine Instruktion für Gruppen 'source monitoring' und 'reality monitoring':*

Liebe Schülerinnen und Schüler des BW Abensberg!

Zuerst einmal vielen Dank, dass Ihr bei uns mitmacht!

Wir wollen herausfinden, wie gut sich Jugendliche Dinge merken können und wie diese Inhalte bei Befragungen erzählt werden.

Wir werden Euch nun eine Bilderserie zeigen, in denen es um einen Verkehrsunfall geht. Es wird bei den Bildern nicht gesprochen. Bitte konzentriert Euch gut auf die Bilder und merkt Euch alles möglichst gut.

Stellt Euch einfach vor, Ihr würdet bei einer Gerichtssendung („Jugendgericht“, „Verkehrsgericht“ usw.) teilnehmen und müsstet dort über das Unfallgeschehen als Zeugen aussagen.

Viel Spaß!

- *Instruktion für Gruppe 'source monitoring' (Befragungszeitpunkte T1 und T2):*

Vielen Dank!

Wir zeigen Euch nun noch einen zweiten Unfall, den Ihr Euch bitte auch aufmerksam anschaut. Denkt daran, dass Ihr auch über diesen Unfall später wie ein Augenzeuge berichten sollt.

In der nächsten Woche werden wir Euch dann zu den beiden Unfällen befragen, außerdem kommen wir nach den Sommerferien noch einmal und werden noch einmal Fragen zu den Bilderserien stellen.

Bitte notiert uns dazu noch die folgenden Informationen:

Name:

Klasse:

Geburtsdatum:

- *Instruktion für Gruppe 'source monitoring' (Befragungszeitpunkt NT2):*

Vielen Dank!

Wir zeigen Euch nun noch einen zweiten Unfall, den Ihr Euch bitte auch aufmerksam anschaut. Denkt daran, dass Ihr auch über diesen Unfall später wie ein Augenzeuge berichten sollt.

Zeugen müssen bei tatsächlichen Gerichtsverhandlungen oft erst mehrere Wochen, nachdem sie etwas beobachtet haben, aussagen. Aus diesem Grund werden wir Euch erst nach den Sommerferien zu den Bilderserien befragen.

Bitte notiert uns dazu noch die folgenden Informationen:

Name:

Klasse:

Geburtsdatum:

- *Instruktion und auszufüllender Bogen für Gruppe 'reality monitoring' bezüglich selbst ausgedachter Unfall-Serie:*

Vielen Dank!

Auf Euch wartet nun folgende Aufgabe: Ihr sollt Euch nun selbst noch einen anderen Unfall ausdenken und diesen dann auch bei den kommenden Befragungen erzählen.

Bitte geht dabei folgendermaßen vor:

- Stellt Euch nun noch einmal die Straßenkreuzung vor, an der sich der erste Unfall ereignet hat.
- Euer „erfundener“ Unfall soll an derselben Kreuzung spielen. Ihr sollt aber frei erfinden, wer aus welcher Richtung kommt, wer schuld ist und wie sich der Unfall überhaupt abspielt (also bitte nicht einfach die Situation aus dem ersten Film übernehmen)
- Der Unfall soll sich zwischen einem Fahrradfahrer (männlich) und einer Autofahrerin (weiblich) ereignen
- Zwei Frauen kommen wieder zufällig hinzu und beobachten das Geschehen
- Der Unfall soll möglichst realistisch sein, Ihr sollt darüber so berichten, wie wenn Ihr ihn tatsächlich gesehen hättet

Bitte beschreibt nun möglichst ausführlich den Unfall, den Ihr Euch selbst ausgedacht habt (gerne auch mit Skizze) ggf. auf Rückseite weiterschreiben:

...

Bitte beschreibt nun noch die Personen, die ihr Euch ausgedacht habt:

	Haare (Farbe, Länge, Frisur)	Statur / Größe	Kleidung (Kleidungsstücke/ Schnitt / Farbe?)	Sonstiges (Bart? Brille?)
Fahrradfahrer (männlich)				
Autofahrerin (weiblich)				
Zeugin 1				
Zeugin 2				

Bitte beschreibt nun noch die Fahrzeuge, die an dem Unfall beteiligt waren:

Fahrrad:

PKW (Marke, Farbe etc.):

- *Instruktion für Gruppe 'reality monitoring' (Befragungszeitpunkte T1 und T2):*

Vielen Dank!

In der nächsten Woche werden wir Euch dann zu den beiden Unfällen befragen, außerdem kommen wir nach den Sommerferien noch einmal und werden noch einmal Fragen zu der Bilderserie und eurer eigenen Unfall-Geschichte stellen.

Bitte notiert uns dazu noch die folgenden Informationen:

Name:

Klasse:

Geburtsdatum:

- *Instruktion für Gruppe 'reality monitoring' (Befragungszeitpunkt NT2):*

Vielen Dank!

Zeugen müssen bei tatsächlichen Gerichtsverhandlungen oft erst mehrere Wochen, nachdem sie etwas beobachtet haben, aussagen. Aus diesem Grund werden wir Euch erst nach den Sommerferien zu den Unfällen befragen.

Bitte notiert uns dazu noch die folgenden Informationen:

Name:

Klasse:

Geburtsdatum:

Anhang B: Interviewleitfaden**Begrüßung, Kontakt und Rapport, Zweck des Interviews** (in etwa so)

„Hallo, mein Name ist Berger von der Uni Regensburg. Du erinnerst dich ja sicher noch (bzw. wahrscheinlich noch) an letzte Woche (bzw. an das letzte Mal bzw. an vor sieben Wochen). Da war ich doch bei euch und hab euch so Dia-Filme gezeigt (bzw. einen Dia-Film und ihr solltet euch dann danach auch selber noch einen weiteren Unfall ausdenken und den aufschreiben...). Es ging ja darum zu schauen, was Ihr euch von diesen Filmen merken konntet und deswegen bin ich jetzt da und möchte dich dazu befragen, was in den Filmen bzw. in deiner Unfallgeschichte so alles vorgekommen ist.“

Freier Bericht

„Gut, und jetzt erzähl mir einfach mal, um was es in den Filmen denn gegangen ist oder alles was du noch weißt dazu“

- „Erzähl mir doch mal, wie das war, um was es gegangen ist!“
- „Erzähl doch mal was Du da beobachtet hast!“

{Aktives Zuhören!}

Möglichkeiten, um den freien Bericht noch weiter anzuregen:

- „ Gut. Fällt Dir sonst noch was ein?“
 - „Kannst Du mir das mit dem noch mal genauer erzählen?“
 - „Und wie war das genau?“
 - „Und sonst?“
 - „Das hab ich jetzt nicht so genau verstanden, kannst Du mir das noch mal genauer erzählen?“
-

Fragenteil

{Weitere Informationen sammeln}

{Wichtig: von offenen Fragen zu geschlossenen Fragen!}

{Fragenteil am besten getrennt für jeden Film durchgehen}

„ Gut, sich stell Dir jetzt noch ein paar Fragen, die mich noch ein bisschen genauer interessieren.“

Zu Personen:

- „Welche Personen waren dabei?“
- „Was haben die Personen denn gemacht?“
- „Wie haben die Personen ausgesehen?“
- „Kannst du mir bitte die Personen beschreiben, die da beteiligt waren?“
- „Wie sah Radfahrer/ Oma/Autofahrer... aus?“
- „Ist Dir an den Personen was Besonderes aufgefallen?“
- „Was hatten die an?“
- „Welche Farbe hatte denn die Hose, Jacke etc.?“
- „Welche Statur hatten die denn?“
- „Waren die dick, dünn oder mittel?“
- „ Wie groß waren die denn?“
- „Waren die groß, klein oder mittel?“
- „Welche Frisuren hatten die denn?“
- „Hatten die lange, kurze oder mittellange Haare?“
- „Sonst irgendwas Besonderes? Vielleicht ein Bart, eine Brille oder so etwas?“
- „Hat sich jemand verletzt?“
- „Wo waren die Personen jeweils gestanden oder gegangen?“ (Position?)

Zu Fahrzeugen: *(incl. Kinderwagen, falls er erwähnt wird)*

- „Welche Fahrzeuge waren denn da dabei?“
- „Weißt Du noch wie das Auto/Fahrrad... ausgesehen hat?“
- „Was war denn das für ein Auto?“
- „Kannst du mir sagen, welche Marke das ... hatte?“

- „Wie groß war denn das Auto? Klein mittel oder groß? Oder wie?“
- „Was hatte das ... für eine Farbe?“
- „War an dem ... denn was kaputt?“

Zu Örtlichkeiten:

- „Kannst Du mir die Umgebung beschreiben, wo das ganze passiert ist?“
- „Wie sah es denn dort aus?“
- „Wo ist denn das passiert?“

Zuordnungsfragen: (bei groben Unklarheiten)

- „Was haben die Personen gemacht?“
- „Welche Fahrzeuge waren denn genau wo beteiligt, das hab ich jetzt nicht so genau verstanden?“
- „Welche Personen waren denn genau wo beteiligt?“
- „Waren da noch andere Leute dabei?“
- „Waren da Zeugen dabei bei dem Unfall?“
- „Wer war denn noch dabei? Woher kam der?“
- „Wer hat denn genau das Auto/Fahrrad... gefahren?“

Frage nach Schuld

- „Wer glaubst du, war da schuld?“
-

Zusammenfassung

„Ich fass das jetzt noch mal kurz zusammen was Du mir erzählt hast, damit Du siehst ob ich alles richtig verstanden habe! Wenn Dir noch was einfällt oder wenn was nicht stimmt, unterbrich mich ruhig.“

„Hat das jetzt so gestimmt?“

Fällt Dir noch was Wichtiges ein, das Du mir bisher nicht erzählt hast?

Schlussphase

{Bedanken für Mitarbeit + verabschieden}

- **Trichterförmig fragen (Offene Fragen → Geschlossene Fragen)**
Wahlfragen (oder wie?)
Leerfrage (wer?, wo? Was?, wie?, womit?)
- **Aktives Zuhören „Mmmmm“**
- **Jugendlichen mit Namen ansprechen**
- **Beim freien Bericht mitnotieren, zwecks Fragenteil und Zusammenfassung!**

- Gruppe 'reality monitoring':

EXTERNAL/INTERNAL**VP : R.T1 / T2 / NT2.G2.X**

Für **DIASERIE 1 – 'PKW-PKW'** und **SERIE 2 – selbst ausgedachter Unfall (analog zu Dia-Serie 'PKW-Fahrrad')**

1. Korrekte Details

a) aus free-recall: ... ⇒ ... Details

b) auf Nachfrage: ... ⇒ ... Details

Gesamt ⇒ ... Details

2. Fehler

a) Quellenverwechslungen: ... ⇒ ... QVen
(innerhalb, zwischen und gesamt)

b) schematypische Intrusionen: ... ⇒ ... Intrusionen

c)Konfabulationen: ... ⇒ ... Konfabulationen

d) allgemeine Erinnerungsfehler: ... ⇒ ... AEF
(errors of commission)

Gesamt ⇒ ... Fehler

Anhang D 1: Stimulusmaterial - Dia-Serie 'PKW-PKW'

Vorführung beginnt!

Dia-Serie 'PKW-PKW' - Szene 1



Dia-Serie 'PKW-PKW' - Szene 2



Dia-Serie 'PKW-PKW' - Szene 3



Dia-Serie 'PKW-PKW' - Szene 4



Dia-Serie 'PKW-PKW' - Szene 5



Dia-Serie 'PKW-PKW' - Szene 6



Dia-Serie 'PKW-PKW' - Szene 7



Dia-Serie 'PKW-PKW' - Szene 8



Dia-Serie 'PKW-PKW' - Szene 9



Dia-Serie 'PKW-PKW' - Szene 10



Dia-Serie 'PKW-PKW' - Szene 11



Dia-Serie 'PKW-PKW' - Szene 12



Dia-Serie 'PKW-PKW' - Szene 13



Dia-Serie 'PKW-PKW' - Szene 14



Dia-Serie 'PKW-PKW' - Szene 15



Dia-Serie 'PKW-PKW' - Szene 16



Dia-Serie 'PKW-PKW' - Szene 17



Dia-Serie 'PKW-PKW' - Szene 18



Dia-Serie 'PKW-PKW' - Szene 19



Dia-Serie 'PKW-PKW' - Szene 20

Vorführung zu Ende!

Dia-Serie 'PKW-PKW' - Szene 21

Anhang D 2: Stimulusmaterial - Dia-Serie 'PKW-Fahrrad'

Vorführung beginnt!

Dia-Serie 'PKW-Fahrrad' - Szene 1



Dia-Serie 'PKW-Fahrrad' - Szene 2



Dia-Serie 'PKW-Fahrrad' - Szene 3



Dia-Serie 'PKW-Fahrrad' - Szene 4



Dia-Serie 'PKW-Fahrrad' - Szene 5



Dia-Serie 'PKW-Fahrrad' - Szene 6



Dia-Serie 'PKW-Fahrrad' - Szene 7



Dia-Serie 'PKW-Fahrrad' - Szene 8



Dia-Serie 'PKW-Fahrrad' - Szene 9



Dia-Serie 'PKW-Fahrrad' - Szene 10



Dia-Serie 'PKW-Fahrrad' - Szene 11



Dia-Serie 'PKW-Fahrrad' - Szene 12



Dia-Serie 'PKW-Fahrrad' - Szene 13



Dia-Serie 'PKW-Fahrrad' - Szene 14



Dia-Serie 'PKW-Fahrrad' - Szene 15



Dia-Serie 'PKW-Fahrrad' - Szene 16



Dia-Serie 'PKW-Fahrrad' - Szene 17



Dia-Serie 'PKW-Fahrrad' - Szene 18



Dia-Serie 'PKW-Fahrrad' - Szene 19



Dia-Serie 'PKW-Fahrrad' - Szene 20

Vorführung zu Ende!

Dia-Serie 'PKW-Fahrrad' - Szene 21

Anhang E: Themenliste der Glaubhaftigkeitsuntersuchung

Tabelle A: Fragestellung 'Glaubhaftigkeit' - Themen der Versuchspersonen (erlebnis-fundiert bzw. nicht erlebnis-fundiert)

Versuchsperson	erlebnis-fundiert	nicht erlebnis-fundiert
1	Arm / Bein gebrochen	Sturz
2	Sturz	Schlägerei mit Gleichaltrigen
3	Schlägerei	Operation
4	Arztbesuch	Verkehrsunfall
5	mit Messer geschnitten	Hundebiss
6	Hundebiss	Sturz / Arm gebrochen
7	Verkehrsunfall	Straßenbahnunfall
8	Sturz	Verkehrsunfall
9	Hundebiss	Bein gebrochen
10	Operation	Unfall beim Dachausbau
11	Sturz	bestohlen werden
12	Sturz	Verkehrsunfall
13	Arm / Bein gebrochen	Sturz
14	Hundebiss	Verkehrsunfall
15	Arztbesuch	Arm / Bein gebrochen
16	Verletzung am Auge	Hundebiss
17	Zahnarztbesuch	Arm gebrochen
18	Arm / Bein gebrochen	Verkehrsunfall
19	Verkehrsunfall	Sturz
20	Brandverletzung	Hundebiss
21	Hundebiss	Sturz
22	Schlägerei	Sturz
23	Fuß verstaucht	Brandverletzung
24	Verkehrsunfall	Hundebiss
25	Bein gebrochen	Schlägerei

Anhang F: allgemeine Instruktion der Glaubhaftigkeitsuntersuchung**Liebe Jugendliche des B.B.W !**

Zuerst einmal vielen Dank, dass Ihr auch bei diesem Projekt mitmacht! Jeder von Euch soll nun zwei Geschichten erzählen, wobei Ihr eine Geschichte wirklich selbst erlebt habt. Die andere Geschichte sollt Ihr Euch ausdenken.

Ich habe hier nun eine Liste mit Themenvorschlägen zusammengestellt.

Bitte kreuze zunächst einmal an, welche dieser Situationen Du schon einmal erlebt hast und wie alt Du damals ungefähr warst:

Thema	Habe ich noch nicht erlebt	Habe ich schon einmal erlebt	Wie alt warst Du damals ungefähr?
von einem Hund/Tier gebissen und verletzt werden			
sich einen Arm oder ein Bein brechen/ verstauchen			
einen Sturz aus größerer Höhe erleiden			
eine Operation im Krankenhaus			
einen Arztbesuch mit Spritze oder Blut abnehmen			
Zahnarztbesuch mit Bohren oder Zahn ziehen			
Prügelei mit Gleichaltrigen			
einen Verkehrsunfall erleiden			
sich mit Messer oder Schere schneiden			
barfuss in etwas (Biene, Scherbe) hineintreten und sich dabei verletzen			

Bitte überlege Dir nun, welche der Situationen, die Du tatsächlich schon erlebt hast, Du mir später kurz und dann in einer Woche dem Hr. Berger erzählen willst.

Als nächstes überlege Dir bitte, zu welchem Thema Du am besten eine Geschichte erfinden könntest.

Ich gehe dann später gleich mit jedem von Euch einzeln die Geschichten kurz durch und wir besprechen dann alles noch mal.

In der nächsten Woche kommt dann der Herr Berger vorbei, dem Ihr die beiden Geschichten erzählen sollt. Er weiß nicht, welche Geschichte Ihr erlebt habt und welche nicht, das soll er selbst heraus bekommen. Deswegen soll auch Eure erfundene Geschichte so klingen, als ob Ihr so etwas wirklich schon erlebt hättet. Achtet darauf, dass Ihr Eure Geschichten in der Ich-Form erzählt und dass sie realistisch sind, weil man sonst sofort merkt, was ausgedacht ist und was nicht. Dazu noch einige Tipps:

- 1) Erzählt immer in der Ich-Form
- 2) Überlegt Euch einen sinnvollen Ablauf der Geschichte
- 3) Wo könnte sich Eure erfundene Geschichte abspielen?
- 4) Wer außer Euch könnte noch beteiligt sein?

Wichtig ist auch, dass Ihr bis zum Ende unseres Projekts nicht untereinander über die Geschichten sprecht.

Vielen Dank!

Anhang G: Interviewleitfaden für Glaubhaftigkeitsuntersuchung**Begrüßung, Kontakt, Rapport, Zweck des Interviews**

Du weißt doch, ich bin heute da, um Dich über die Geschichten zu befragen, die Du Dir überlegt hast.

Die Frau Dr. Loohs war ja letzte Woche schon bei Dir und hat die Geschichten mit Dir besprochen - und auch euer Lehrer. Ich war ja nicht dabei und deshalb weiß ich auch nicht, welche Geschichte du erlebt hast und welche nicht. Das sollst du auch bitte für dich behalten – das wäre ganz wichtig! O.k?

Du sollst mir einfach nur die beiden Geschichten erzählen. Ich weiß das die eine Geschichte davon handelt, wie du dir ein _____ hast
und die andere Geschichte davon handelt, wie du dich _____ hast.

Gut, fangen wir mal an. Magst du mit zuerst die Geschichte von dem _____ erzählen oder die Geschichte vom _____ erzählen?

Eigentlich darf ich ja gar nicht wissen, welche Geschichte du wirklich erlebt hast und welche nicht. Erzähl mir doch einfach die Geschichte so als ob du sie wirklich erlebt hättest.

Verständnis

Ich hatte noch nie/bin noch nie/hab noch nie/ wurde noch nie/,
deshalb kann ich mir das nicht so genau vorstellen. Kannst du mir das nochmals erzählen.

Ich hab das jetzt noch nicht so gut verstanden. Kannst Du mir das noch mal versuchen zu erklären?

Ich war jetzt grad a bissl langsam/ bin heut a bissl müde. Kannst Du mir das vielleicht/bitte noch mal erklären?

Und dann?

Wie war das genau? Ich hab grad Schwierigkeiten, mir das so vorzustellen.

Weißt du, ich hab mir noch nie z.B. ein Bein gebrochen....

Geschichte 1 (analog Geschichte 2)

Freier Bericht

Gut, jetzt erzähl mir einfach mal, wie das war!

- Erzähl mir doch mal was da passiert!

Möglichkeiten um den freien Bericht nochmals anzuregen:

- Super, fällt dir sonst noch was ein?

- Und sonst?

Das was du bis jetzt erzählt hast, war ja schon ganz toll. Aber ich würde mir das ganze gerne noch etwas genauer vorstellen können. Deswegen frag ich Dich noch ein bisschen was:

Fragenteil

Kontext:

Wo?

Wann?

Wie?

- Wie....., wo warst du denn da?

- Was hast du denn da vorher gemacht?

- Und was hast du nachher gemacht?

- Wie kam es denn da genau dazu?

Interaktionen:

Was hast du denn dann gemacht?

Und was haben die anderen gemacht?

- War da eigentlich noch jemand dabei?

Gespräche

Hat da jemand was gesagt?

- Ist da was gesprochen worden?

Gefühle

Wie ging es dir denn da dabei?

Hast du da was gespürt?

- Wie hast du dich denn da gefühlt?

Details

Wie alt warst du denn da, als das passiert ist?

Zu welcher Jahreszeit war denn das?

Was hattest du denn an dem Tag an?

Von wem wurde...?

Was hatten die anderen an?

Anhang H: Ratingbogen**Vp-Code:** _____

Die Merkmale sollen nach folgendem Beurteilungsschema geratet werden:

<i>nicht vorhanden</i>	<i>schwach vorhanden</i>	<i>mittel vorhanden</i>	<i>stark vorhanden</i>	<i>sehr stark vorhanden</i>
0	1	2	3	4

A: Allgemeine Merkmale

	0	1	2	3	4
1. Logische Konsistenz					
Aussage in sich stimmig und ohne Widersprüche?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Widerspruchsfreie ungeordnete Darstellung					
Sprunghafte, unzusammenhängende, nicht chronologische Darstellung ohne Widersprüche?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Quantitativer Detailreichtum					
Wie viele Einzelheiten enthält die Aussage?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

B: Einzelne Inhalte

	0	1	2	3	4
4. Kontextuelle Einbettung					
Zeitlich/örtliche Einbettung des Aussageinhalts?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Beschreibung nonverbaler Aktionen					
Werden nonverbale Aktionen geschildert?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Wiedergabe von Gesprächen					
Werden konkrete Inhalte stattgefundenener Gespräche in der Aussage wiedergegeben?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
7. Komplikationen im Handlungsverlauf					
Enthält die Aussage Schilderungen unerwarteter Komplikationen im dargestellten Handlungsverlauf?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Ausgefallene Einzelheiten					
Kommen in der Aussage Details mit einer geringen Auftretenswahrscheinlichkeit vor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Nebensächliche (überflüssige) Einzelheiten					
Kommen in der Aussage Details vor, die mit dem Tathergang selbst nichts zu tun haben?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Phänomengemäße Schilderung unverstandener Handlungselemente					
Schildert die Person unverstandene Geschehensabläufe?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Inhaltliche Verschachtelungen					
Sind verschiedene Handlungen ineinander verschachtelt?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Schilderung eigenpsychischer Vorgänge					
Beschreibt die aussagende Person (Opfer) eigene psychische oder physische Vorgänge?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Schilderung psychischer Vorgänge beim Täter					
Macht die aussagende Person (Opfer) Beschreibungen über psychische oder physische Vorgänge beim Täter?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. Spontane Verbesserungen der eigenen Aussage					
Verbessert oder präzisiert die aussagende Person spontan ihre inhaltliche Darstellung?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

ABBILDUNGS- UND TABELLENVERZEICHNIS

Abbildungen

Bezeichnung	(verkürzter) Titel bzw. Inhalt	Seite
A-3-1	Verteilung der IQ-Werte (source- und reality monitoring)	148
A-3-2	Verteilung der IQ-Werte (Glaubhaftigkeit)	162
A-4-1	Gruppe 'ESM' und 'RM': korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler 'gesamt' zu T1	179
A-4-2	Gruppen 'ESM' und 'RM': Quellenverwechslung 'gesamt', 'innerhalb' und 'zwischen', schematypische Intrusionen und Konfabulationen zu T1	180
A-4-3	Gruppen 'ESM' und 'RM': allgemeine Erinnerungsfehler zu T1	181
A-4-4	Gruppe 'ESM' und 'RM': korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler 'gesamt' zu T2	186
A-4-5	Gruppen 'ESM' und 'RM': Quellenverwechslung 'gesamt', 'innerhalb' und 'zwischen', schematypische Intrusionen und Konfabulationen zu T2	188
A-4-6	Gruppen 'ESM' und 'RM': allgemeine Erinnerungsfehler zu T2	189
A-4-7	Gruppe 'ESM' und 'RM': korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler 'gesamt' zu NT2	194
A-4-8	Gruppen 'ESM' und 'RM': Quellenverwechslung 'gesamt', 'innerhalb' und 'zwischen', schematypische Intrusionen und Konfabulationen zu NT2	196
A-4-9	Gruppen 'ESM' und 'RM': allgemeine Erinnerungsfehler zu NT2	197
A-4-10	korrekte Details gesamt und Fehler 'gesamt' für ESM-Gruppe zu T1-T2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	206
A-4-11	Fehlerarten für ESM-Gruppe zu T1-T2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	207
A-4-12	korrekte Details 'gesamt' und Fehler 'gesamt' für RM-Gruppe zu T1-T2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	213
A-4-13	Fehlerarten für RM-Gruppe zu T1-T2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	214
A-4-14	korrekte Details 'gesamt', Fehler 'gesamt' und allgemeine Erinnerungsfehler für ESM-Gruppe zu T1-NT2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	220

A-4-15	Fehlerarten für ESM-Gruppe zu T1-NT2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	221
A-4-16	korrekte Details 'gesamt', Fehler 'gesamt' und allgemeine Erinnerungsfehler für RM-Gruppe zu T1-NT2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	227
A-4-17	Fehlerarten für RM-Gruppe zu T1-NT2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	228
A-4-18	korrekte Details 'gesamt', Fehler 'gesamt' und allgemeine Erinnerungsfehler für ESM-Gruppe zu T2-NT2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	234
A-4-19	Fehlerarten für ESM-Gruppe zu T2-NT2 für beide Dia-Serien 'PKW-PKW' und 'PKW-Fahrrad'	235
A-4-20	korrekte Details 'gesamt', Fehler 'gesamt' und allgemeine Erinnerungsfehler für RM-Gruppe zu T2-NT2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	241
A-4-21	Fehlerarten für RM-Gruppe zu T2-NT2 für Dia-Serie 'PKW-PKW' und selbst erfundene Unfall-Serie	242
A-4-22 (a-c)	Haupt- und Interaktionseffekte für korrekte Details 'gesamt', im 'free recall' und auf 'Nachfrage'	249/ 250
A-4-23 (a-g)	Haupt- und Interaktionseffekte für Fehler 'gesamt' und Fehlerarten	251 - 253
A-4-24-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Logische Konsistenz	260
A-4-24-2	Haupt- und Interaktionseffekte zu T1 und T2: Logische Konsistenz	260
A-4-25-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Widerspruchsfreie, ungeordnete Darstellung	262
A-4-25-2	Haupt- und Interaktionseffekte zu T1 und T2: Widerspruchsfreie, ungeordnete Darstellung	262
A-4-26-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Quantitativer Detailreichtum	264
A-4-26-2	Haupt- und Interaktionseffekte zu T1 und T2: Quantitativer Detailreichtum	264
A-4-27-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Kontextuelle Einbettung	266
A-4-27-2	Haupt- und Interaktionseffekte zu T1 und T2: Kontextuelle Einbettung	266

A-4-28-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung nonverbaler Aktionen	268
A-4-28-2	Haupt- und Interaktionseffekte zu T1 und T2: Schilderung nonverbaler Aktionen	268
A-4-29-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Wiedergabe von Gesprächen	270
A-4-29-2	Haupt- und Interaktionseffekte zu T1 und T2: Wiedergabe von Gesprächen	270
A-4-30-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung von Komplikationen im Handlungsverlauf	272
A-4-30-2	Haupt- und Interaktionseffekte zu T1 und T2: Schilderung von Komplikationen im Handlungsverlauf	272
A-4-31-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Ausgefallene Details	274
A-4-31-2	Haupt- und Interaktionseffekte zu T1 und T2: Ausgefallene Details	274
A-4-32-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Nebensächliche Details	276
A-4-32-2	Haupt- und Interaktionseffekte zu T1 und T2: Nebensächliche Details	276
A-4-33-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung eigener psychischer oder physischer Vorgänge	278
A-4-33-2	Haupt- und Interaktionseffekte zu T1 und T2: Schilderung eigener psychischer oder physischer Vorgänge	278
A-4-34-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	280
A-4-34-2	Haupt- und Interaktionseffekte zu T1 und T2: Schilderung psychischer oder physischer Vorgänge beim Täter oder Beteiligten	280
A-4-35-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Spontane Verbesserungen der eigenen Aussage	282
A-4-35-2	Haupt- und Interaktionseffekte zu T1 und T2: Spontane Verbesserungen der eigenen Aussage	282
A-4-36-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Zugeben von Erinnerungslücken	284

A-4-36-2	Haupt- und Interaktionseffekte zu T1 und T2: Zugeben von Erinnerungslücken	284
A-4-37-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Einwände gegen die Richtigkeit der eigenen Aussage	286
A-4-37-2	Haupt- und Interaktionseffekte zu T1 und T2: Einwände gegen die Richtigkeit der eigenen Aussage	286
A-4-38-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Selbstbelastungen	288
A-4-38-2	Haupt- und Interaktionseffekte zu T1 und T2: Selbstbelastungen	288
A-4-39-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Inschutznahme des Täters oder Beteiligten	290
A-4-39-2	Haupt- und Interaktionseffekte zu T1 und T2: Inschutznahme des Täters oder Beteiligten	290
A-4-40-1	Vergleich der Gruppenmittelwerte zu T1 und T2: Deliktspezifische Aussageelemente	292
A-4-40-2	Haupt- und Interaktionseffekte zu T1 und T2: Deliktspezifische Aussageelemente	292

Tabellen

Nummer	(verkürzter) Titel bzw. Inhalt	Seite
T-1-1	Gedächtnisprozesse und Aussagetypen	7
T-1-2	Differenzierung nach Zeugengruppen und Aussageleistungen	12
T-1-3	Realkennzeichen nach STELLER & KÖHNKEN	59
T-1-4	Einteilung der Intelligenzminderung	84
T-1-5	Kategorien von 'Mental Retardation'	97
T-1-6	Unangemessene bzw. ungeeignete und problematische Fragestrategien	140
T-1-7	Angemessene bzw. geeignete Befragungstechniken	140
T-3-1	IQ-Daten 1 (Untersuchung 'source-' und 'reality monitoring')	148
T-3-2	Experimentalplan (1) - 'source-/ reality monitoring'	152
T-3-3	Geschlechtsspezifische Verteilung der Versuchspersonen auf die Versuchsgruppen ('source-/ reality monitoring')	154
T-3-4	Abhängige Variablen (1) - 'source-/ 'reality-monitoring'	159
T-3-5	IQ-Daten 2 (Untersuchung 'Glaubhaftigkeit')	162
T-3-6	Themenkataloge	164

T-3-7	Experimentalplan (2) - 'Glaubhaftigkeit'	166
T-3-8	Abhängige Variablen (2) - 'Glaubhaftigkeitsuntersuchung'	172
T-3-9	angewandte statistische Verfahren (einzelne Aussagemerkmale)	174
T-3-10	angewandte statistische Verfahren (Merkmalskategorien)	174
T-4-1	Ergebnisse des K-S – Tests (1) für die Gruppe 'ESM'	176
T-4-2	Ergebnisse des K-S – Tests (2) für die Gruppe 'RM'	176
T-4-3	Varianzhomogenität (1) - 'source- / reality – monitoring'	177
T-4-4	MW und SD der beiden Gruppen für korrekte Details ('gesamt', 'free recall' und 'auf Nachfrage') und 'Fehler gesamt' zu T1 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	179
T-4-5	Ergebnisse t – Test (ESM - RM) für korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zu T1 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	180
T-4-6	MW und SD der beiden Gruppen für Fehlerarten zu T1 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	182
T-4-7	Ergebnisse t – Test (ESM - RM) für Fehlerarten zu T1 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	182
T-4-8	MW und SD der beiden Gruppen für alle Variablen zu T1 für Serie 'PKW-PKW'	183
T-4-9	Ergebnisse t – Test (ESM - RM) für alle Variablen zu T1 für Serie 'PKW-PKW'	184
T-4-10	MW und SD der beiden Gruppen für alle Variablen zu T1 für Serie 'PKW-Fahrrad'	185
T-4-11	Ergebnisse t – Test (ESM - RM) für alle Variablen zu T1 für Serie 'PKW-Fahrrad'	186
T-4-12	MW und SD der beiden Gruppen für korrekte Details ('gesamt', 'free recall' und 'auf Nachfrage') und 'Fehler gesamt' zu T2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	187
T-4-13	Ergebnisse t – Test (ESM - RM) für korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zu T2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	187
T-4-14	MW und SD der beiden Gruppen für Fehlerarten zu T2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	189
T-4-15	Ergebnisse t – Test (ESM - RM) für Fehlerarten zu T2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	190
T-4-16	MW und SD der beiden Gruppen für alle Variablen zu T2 für Serie 'PKW-PKW'	191

T-4-17	Ergebnisse t – Test (ESM - RM) für alle Variablen zu T2 für Serie 'PKW-PKW'	191
T-4-18	MW und SD der beiden Gruppen für alle Variablen zu T2 für Serie 'PKW-Fahrrad'	192
T-4-19	Ergebnisse t – Test (ESM - RM) für alle Variablen zu T2 für Serie 'PKW-Fahrrad'	193
T-4-20	MW und SD der beiden Gruppen für korrekte Details ('gesamt', 'free recall' und 'auf Nachfrage') und 'Fehler gesamt' zu NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	195
T-4-21	Ergebnisse t – Test (ESM - RM) für korrekte Details 'gesamt' - 'im free recall' - 'auf Nachfrage' und Fehler gesamt zu NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	195
T-4-22	MW und SD der beiden Gruppen für Fehlerarten zu NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	198
T-4-23	Ergebnisse t – Test (ESM - RM) für Fehlerarten zu NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	198
T-4-24	MW und SD der beiden Gruppen für alle Variablen zu NT2 für Serie 'PKW-PKW'	200
T-4-25	Ergebnisse t – Test (ESM - RM) für alle Variablen zu NT2 für Serie 'PKW-PKW'	200
T-4-26	MW und SD der beiden Gruppen für alle Variablen zu NT2 für Serie 'PKW-Fahrrad'	202
T-4-27	Ergebnisse t – Test (ESM - RM) für alle Variablen zu NT2 für Serie 'PKW-Fahrrad'	202
T-4-28	Zusammenfassung der Ergebnisse der t-Tests für die ESM- und die RM-Gruppe zu T1, T2 und NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	203
T-4-29	Zusammenfassung der Ergebnisse der t-Tests für die ESM- und die RM-Gruppe zu T1, T2 und NT2 für Serie 'PKW-PKW'	203
T-4-30	Zusammenfassung der Ergebnisse der t-Tests für die ESM- und die RM-Gruppe zu T1, T2 und NT2 für Serie 'PKW-Fahrrad'	204
T-4-31	Zusammenfassung der Ergebnisse der t-Tests für die ESM- und die RM-Gruppe zu T1, T2 und NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad', 'PKW-PKW' und 'PKW-Fahrrad'	205
T-4-32	MW und SD für ESM-Gruppe zu T1-T2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	208
T-4-33	Ergebnisse t-Test für ESM-Gruppe zu T1-T2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	209

T-4-34	MW und SD für ESM-Gruppe zu T1-T2 für Dia-Serie 1 - 'PKW-PKW'	210
T-4-35	MW und SD für ESM-Gruppe zu T1-T2 für Dia-Serie 2 - 'PKW-Fahrrad'	211
T-4-36	Zusammenfassung der Ergebnisse t-Tests für ESM-Gruppe zu T1-T2 für beide Dia-Serien	212
T-4-37	MW und SD für RM-Gruppe zu T1-T2 für Dia-Serie 'PKW-PKW'.selbst erfundene Unfall-Serie	215
T-4-38	Ergebnisse t-Test für RM-Gruppe zu T1-T2 für Dia-Serie 'PKW-PKW'.selbst erfundene Unfall-Serie	215
T-4-39	MW und SD für RM-Gruppe zu T1-T2 für Dia-Serie 1 - 'PKW-PKW'	217
T-4-40	MW und SD für RM-Gruppe zu T1-T2 für selbst erfundene Unfall-Serie	218
T-4-41	Zusammenfassung der Ergebnisse t-Tests für RM-Gruppe zu T1-T2 für beide Unfall-Serien	219
T-4-42	MW und SD für ESM-Gruppe zu T1-NT2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	222
T-4-43	Ergebnisse t-Test für ESM-Gruppe zu T1-NT2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	223
T-4-44	MW und SD für ESM-Gruppe zu T1-NT2 für Dia-Serie 1 - 'PKW-PKW'	224
T-4-45	MW und SD für ESM-Gruppe zu T1-NT2 für Dia-Serie 2 - 'PKW-Fahrrad'	225
T-4-46	Zusammenfassung der Ergebnisse t-Tests für ESM-Gruppe zu T1-NT2 für beide Dia-Serien	226
T-4-47	MW und SD für RM-Gruppe zu T1-NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	229
T-4-48	Ergebnisse t-Test für RM-Gruppe zu T1-NT2 für Serien 'PKW-PKW'. 'PKW-Fahrrad'	229
T-4-49	MW und SD für RM-Gruppe zu T1-NT2 für Dia-Serie 1 - 'PKW-PKW'	231
T-4-50	MW und SD für RM-Gruppe zu T1-NT2 für selbst erfundene Unfall-Serie	232
T-4-51	Zusammenfassung der Ergebnisse t-Tests für RM-Gruppe zu T1-NT2 für beide Unfall-Serien	233
T-4-52	MW und SD für ESM-Gruppe zu T2-NT2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	236
T-4-53	Ergebnisse t-Test für ESM-Gruppe zu T2-NT2 für Dia-Serien 'PKW-PKW'. 'PKW-Fahrrad'	237
T-4-54	MW und SD für ESM-Gruppe zu T2-NT2 für Dia-Serie 1 - 'PKW-PKW'	238
T-4-55	MW und SD für ESM-Gruppe zu T2-NT2 für Dia-Serie 2 - 'PKW-Fahrrad'	239
T-4-56	Zusammenfassung der Ergebnisse t-Tests für ESM-Gruppe zu T2-NT2 für beide Dia-Serien	240

T-4-57	MW und SD für RM-Gruppe zu T2-NT2 für beide Unfall-Serien	243
T-4-58	Ergebnisse t-Test für RM-Gruppe zu T2-NT2 für beide Unfall-Serien	243
T-4-59	MW und SD für RM-Gruppe zu T2-NT2 für Dia-Serie 1 - 'PKW-PKW'	245
T-4-60	MW und SD für RM-Gruppe zu T2-NT2 für selbst erfundene Unfall-Serie	246
T-4-61	Zusammenfassung der Ergebnisse t-Tests für RM-Gruppe zu T2-NT2 für beide Unfall-Serien	247
T-4-62	Zusammenfassung der Ergebnisse der t-Tests für die ESM- und die RM-Gruppe zu T1-T2, T1-NT2 und T2-NT2 für Serien 'PKW-PKW', 'PKW-Fahrrad', 'PKW-PKW' und 'PKW-Fahrrad' bzw. selbst erfundene Unfall-Serie (analog 'PKW-Fahrrad')	248
T-4-63	Ergebnisse des ALM – Haupteffekt 'Gruppe'	254
T-4-64	Ergebnisse des ALM – Haupteffekt 'Befragungszeitpunkt'	254
T-4-65	Ergebnisse des ALM – Interaktion 'Gruppe-Befragungszeitpunkt'	255
T-4-66	Effektstärken (1) - 'source-/ reality monitoring'	255
T-4-67	Zusammenfassung der Ergebnisse des ALM (1) - 'source-/ reality monitoring'	256
T-4-68	Ergebnisse des K-S-Tests (3) für alle Aussagemerkmale und Merkmalskategorien	258
T-4-69	Ergebnisse des ALM (2 - GLH)	298
T-4-70	Ergebnisse aus den post-hoc-Verfahren (Aussagemerkmale - einzeln - GLH)	299
T-4-71	Effektstärken (2) – GLH (Aussagemerkmale – einzeln)	300
T-4-72	Ergebnisse des ALM (3 - GLH)	301
T-4-73	Ergebnisse aus den post-hoc-Verfahren (Kategorien der Aussagemerkmale und Aussagemerkmale 'total' - GLH)	301
T-4-74	Effektstärken (3) – GLH (Kategorien der Aussagemerkmale und Aussagemerkmale 'total')	302
T-4-75	Glaubhaftigkeitseinschätzung (T1 und T2)	303
T-4-76	Glaubhaftigkeitseinschätzung (T1)	304
T-4-77	Glaubhaftigkeitseinschätzung (T2)	305
T-4-78	Gleichheitstest der Gruppenmittelwerte zu T1	307
T-4-79	Gleichheitstest der Gruppenmittelwerte zu T2	309
T-4-80	Standardisierte kanonische Diskriminanzfunktionskoeffizienten	311
T-4-81	Rangfolge der gewichteten diskriminatorisch bedeutsamen Aussagemerkmale zu beiden Befragungszeitpunkten T1 und T2	312
T-4-82	Statistische Signifikanzprüfung der Diskriminanzfunktionen für T1 und T2	313

T-4-83	Gruppenmittelwerte und kritische Diskriminanzwerte für die Diskriminanzgruppen zu beiden Befragungszeitpunkten T1 und T2	314
T-4-84	Klassifikationsmatrix (Klassifizierungsergebnisse) zu T1	315
T-4-85	Klassifikationsmatrix (Klassifizierungsergebnisse) zu T2	316
Tabelle A	Fragestellung 'Glaubhaftigkeit' - Themen der Versuchspersonen	477

EIDESSTATTLICHE VERSICHERUNG

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter der Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

- Frau Dr. Sandra Loohs:

Diskussion, Auswahl und zur Verfügung Stellen der als Stimulusmaterial verwendeten Dia-Serien

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir mittelbar oder unmittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Vor Aufnahme der obigen Versicherung an Eides Statt wurde ich über die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung belehrt.

Regensburg, 09.05.2005

(Oliver Berger)

LEBENS LAUF

Oliver Adrian Berger

Persönliche Angaben:

Name: Oliver Adrian Berger
Anschrift: Am Protzenweiher 17A - 93059 Regensburg
Geburtstag: 27.06.1974
Geburtsort: Passau
Eltern: Helene und Wolfgang Berger, geschieden

Schulbildung:

Sept. 1981 bis Juli 1985: Besuch der Grundschule St. Nikola - Passau
Sept. 1985 bis Juli 1994: Besuch des Gymnasiums Leopoldinum - Passau
(humanistisches und neusprachliches Gymnasium)
Abschluss am 01.07.1994: Allgemeine Hochschulreife (Abitur)

Studium:

WS 1994/95 bis SS 1995: Studium des Lehramts für Grundschulen an der Universität Regensburg
WS 1995/96 bis SS 2002: Studium der Psychologie an der Universität Regensburg
- Diplom – Vorprüfung nach dem SS 1998
- 3 psychologische Praktika in den Bereichen `Diagnostik`,
`Kinder- und Jugendpsychiatrie`, `Sucht` (SS 98 bis WS 99)
- Tätigkeit als Sitzwache auf den Stationen der Fachklinik für
Kinder- u. Jugendpsychiatrie des BKR (SS '99 bis SS '02)
- Diplom – Prüfung nach dem SS 2002 (Gesamtnote 1,1)

Promotion:

Ab März 2003 erhielt ich im Rahmen der Graduiertenförderung der Universität Regensburg ein Promotionsstipendium, in deren Rahmen wiederum die vorliegende Dissertation entstand.